

Improving Remote Sensed Data Products Using Bayesian Methodology for the Analysis of Computer Model Output

Grant Number NNX07AV69G – Second Year Report

Robin D. Morris, USRA-RIACS
Athanasios Kottas, UCSC
Roberto Furfaro, Barry D. Ganapol, UofA

1st August 2009

Abstract

We report on progress made during the second year of this grant. The work performed in the second year has built upon the work described in last year's report, and provides the foundation for the validation and inversion work to be performed in year three.

The main items in year two are

1. The completion of the extension of the Leaf Canopy Model (LCM) to model the full bi-directional reflection distribution function, and new methodology for solving the radiation transport equations.
2. The extraction of informative prior distributions over the inputs of the LCM from the LOPEX database.
3. The completion of the development of the fully-Bayesian inference approach to sensitivity analysis.
4. The completion of a major review of the retrieval of vegetation canopy biophysical parameters from remote sensing data and radiative transfer models.
5. The award of a Outreach Supplement, which will be used to produce a short video for broadcast on PBS and distribution via YouTube.

Much of this work has been either submitted for publication or is in an advanced stage of preparation. Preprints are attached to this report.

We conclude with plans for year three, and a list of presentations and publications.

1 Completion of the extension of the LCM to the full BRDF and new methodology for solving the radiation transport equations

The extension of the LCM to compute the bi-directional reflection distribution function has been completed. This is an important step towards the goal of model inversion, as it allows the prediction of data taking into account the time, date and place of measurement (which determines the sun angles at the area observed), and also the viewing geometry of the satellite. This work is considered to be an extension of the LCM model, and so has not been submitted for publication independently.

	band; wavelength (nm)							
input	1 469	2 555	3 1240	4 1640	5 2130	6 667	7 748	8 870
LAI	0.07	0.03	0.60	0.38	0.15	0.27	0.44	0.66
CHL	0.50	0.71	0.00	0.00	0.00	0.31	0.19	0.00
Water	0.00	0.00	0.01	0.06	0.30	0.00	0.00	0.00
Protein	0.00	0.00	0.01	0.00	0.00	0.00	0.02	0.02
Lignin	0.00	0.01	0.08	0.18	0.21	0.00	0.14	0.06
Thickness	0.02	0.02	0.13	0.12	0.03	0.02	0.06	0.11
Soil	0.00	0.00	0.03	0.03	0.01	0.00	0.02	0.04
Sensor azimuth	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Sensor zenith	0.00	0.01	0.01	0.01	0.01	0.01	0.00	0.00
Sun azimuth	0.00	0.00	0.00	0.01	0.02	0.00	0.00	0.00
Sun zenith	0.00	0.00	0.00	0.00	0.01	0.00	0.00	0.00
Total	0.60	0.78	0.86	0.78	0.73	0.62	0.87	0.90

Table 1: Sensitivity Indices for bi-directional extension of LCM; Planophile canopy.

It will be included as part of an upcoming paper describing certain extensions to the statistical analysis framework that we published last year in the IEEE Transactions on Geoscience and Remote Sensing.

Work on a new methodology for solving the radiative transport equations for vegetation, the Analytical Discrete Ordinate method, has been presented at the International Conference on Transport Theory, and is currently being prepared for publication.

In figure 1 we show the main effects for the bi-directional extension of the LCM for a planophile canopy. There are some changes in the shapes of the main effects for the LCM inputs considered so far (e.g. LAI in band 5 and CHL in band 6), but the relative importance of each input has not changed. The main effects for the newly considered inputs, the sun and sensor angles, are shown in the lower panels. Note that these are an order of magnitude smaller than the main effects for the other inputs, for the planophile case. Table 1 confirms this result, where the sensitivity indices for the sun and sensor angles are all very small. Note, however, that the total of the sensitivity indices are now much reduced, indicating that there may be important interaction effects between the sun and sensor angles and the other inputs.

In figure 2 we show the main effects for an Erectophile canopy. In this case, for some of the spectral bands the main effects for the sun and sensor angles are of the same order of magnitude as the main effects for the other inputs. This is confirmed by the sensitivity indices given in table 2, where the sensitivity indices for the angular inputs are now comparable with the others. This stresses the importance of considering the full bi-directional distribution, and also the importance of the categorical Leaf Angle Distribution input to reliable inversion.

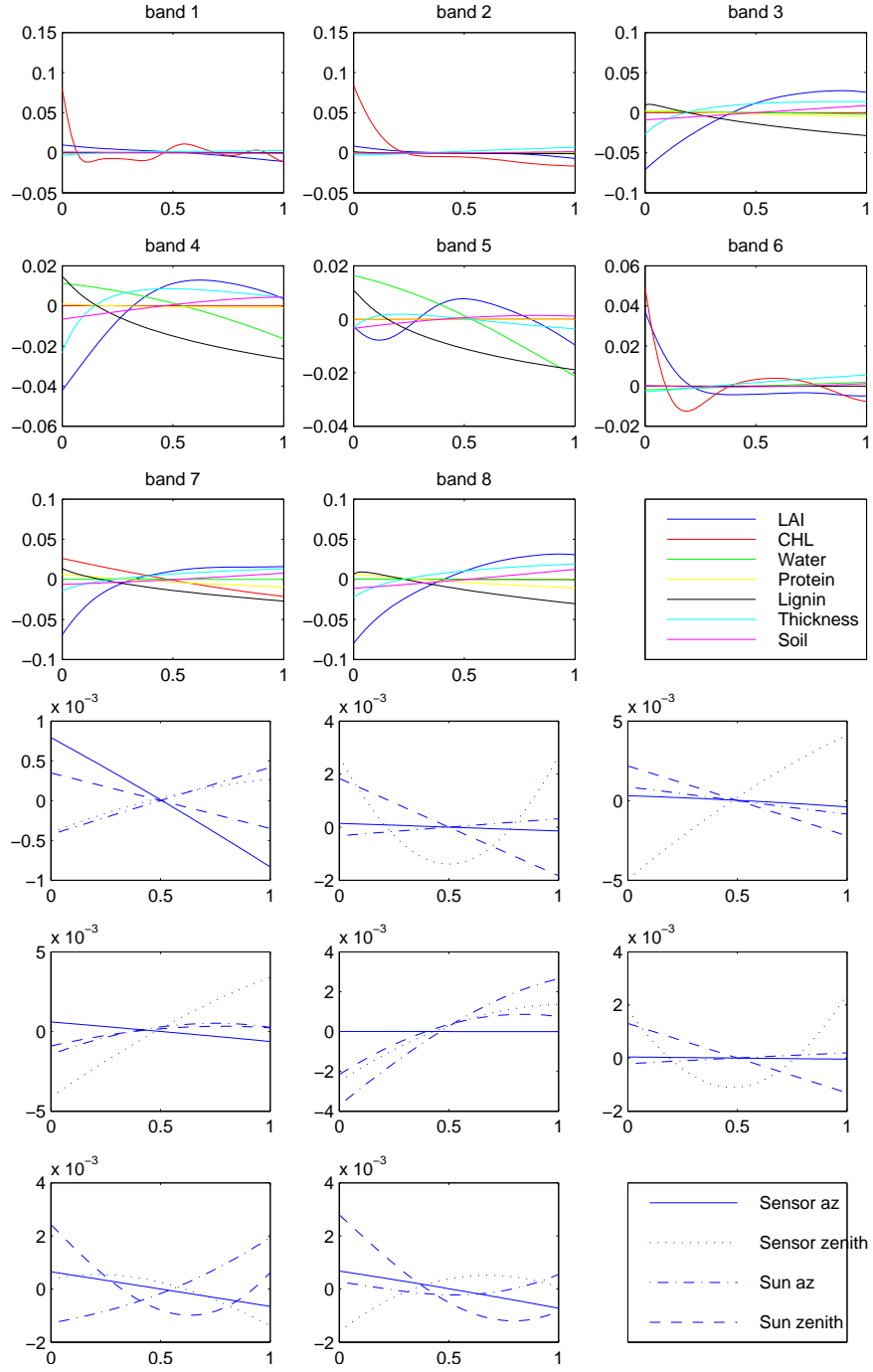


Figure 1: Main Effects for bi-directional extension of LCM; Planophile canopy.

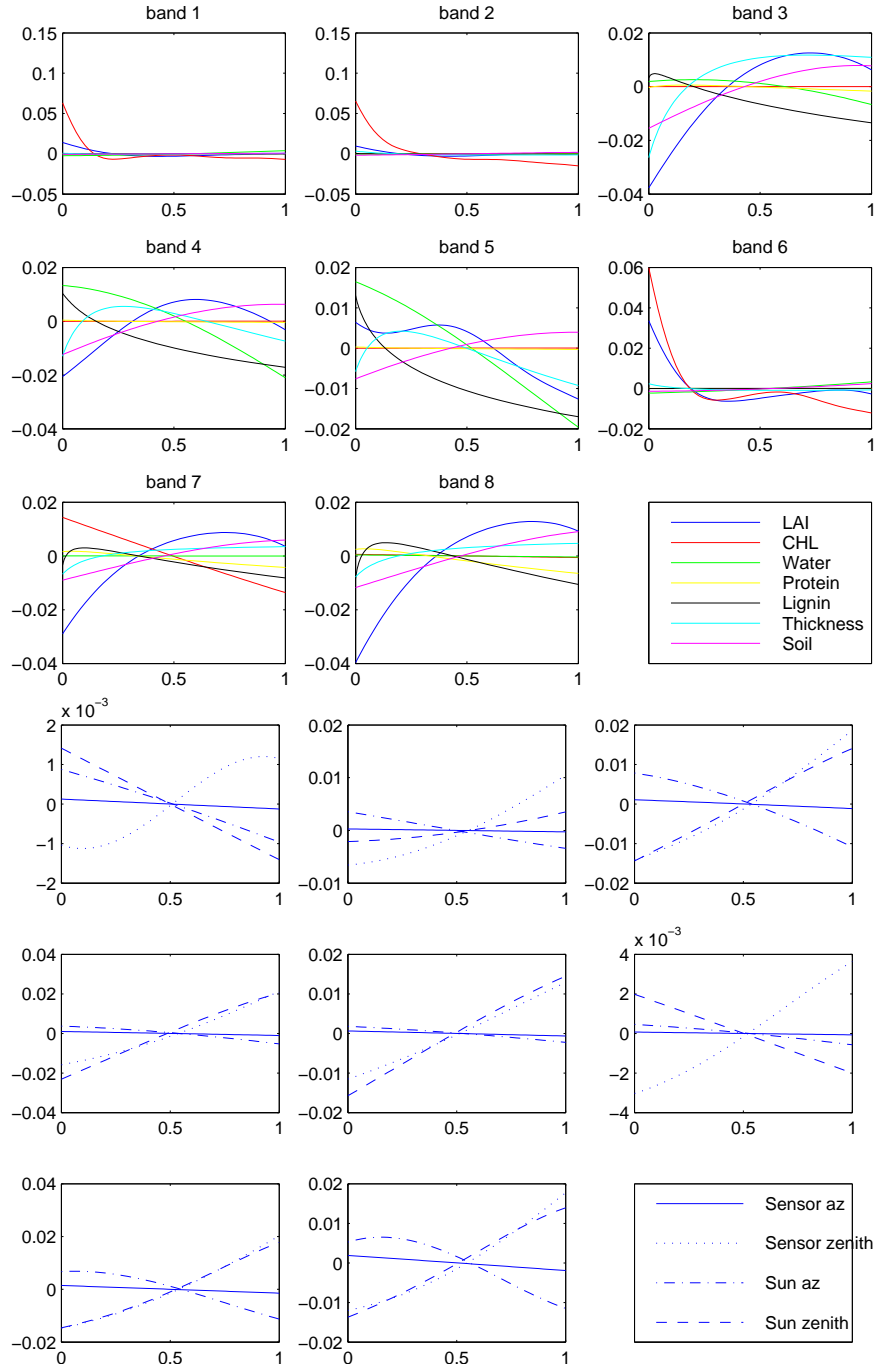


Figure 2: Main Effects for bi-directional extension of LCM; Erectophile canopy.

	band; wavelength (<i>nm</i>)							
input	1 469	2 555	3 1240	4 1640	5 2130	6 667	7 748	8 870
LAI	0.06	0.03	0.33	0.14	0.12	0.17	0.18	0.30
CHL	0.59	0.56	0.12	0.00	0.00	0.49	0.07	0.00
Water	0.02	0.00	0.03	0.11	0.17	0.01	0.00	0.00
Protein	0.00	0.00	0.00	0.00	0.00	0.00	0.01	0.01
Lignin	0.00	0.00	0.04	0.08	0.15	0.00	0.03	0.05
Thickness	0.01	0.02	0.21	0.07	0.09	0.01	0.05	0.04
Soil	0.00	0.01	0.08	0.05	0.04	0.01	0.05	0.07
Sensor azimuth	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Sensor zenith	0.01	0.07	0.11	0.13	0.09	0.02	0.14	0.09
Sun azimuth	0.00	0.01	0.04	0.02	0.01	0.00	0.07	0.06
Sun zenith	0.00	0.01	0.09	0.17	0.14	0.00	0.13	0.08
Total	0.69	0.71	0.81	0.78	0.79	0.70	0.73	0.70

Table 2: Sensitivity Indices for bi-directional extension of LCM; Erectophile canopy.

2 Extraction of informative prior distributions over the inputs of the LCM from the LOPEX database

The LOPEX database is a set of measurements of leaves and plants made by scientists at the Joint Research Center – Institute of Remote Sensing Applications in Italy. Their primary interest was in the spectroscopic response of the leaves, and indeed, the LOPEX database is dominated by spectral information. They did also, however, measure a great deal of auxiliary information about the leaves that they analyzed, some of which provides useful prior information regarding the expected values of the inputs to the LCM. For example, figure 3 shows the distribution of water fraction for each of the leaf samples for which this data is available. The variability within each tree type is clearly visible in the box plots for each sample number. The data can be aggregated to give a distribution over water fraction for all the trees in the database. Figure 4 shows the prior for water fraction (in green) and the other inputs for which there is information in the LOPEX database. (Note that these distributions are rescaled on to a 0-1 range for ease of display.) The LOPEX database can be seen to provide a great deal of information about some of the inputs to the LCM – the distributions shown on figure 4 are informative compared with the uniform distributions we assumed up to this point.

To see the effects of these priors, figures 5 and 6 show the main effects, and tables 3 and 4 show the sensitivity indices, with and without the LOPEX priors. Consider the sensitivity indices in band 1 for LAI and CHL. The LOPEX priors provide significant information about the distribution of CHL expected – this causes the sensitivity index for CHL to be significantly reduced (from 0.74 to 0.39), and the sensitivity index for LAI to be significantly increased (from 0.05 to 0.38). A similar effect is seen in band 5, where the information in the LOPEX prior for the water fraction has reduced the sensitivity index for water and caused a corresponding increase in the sensitivity index for LAI. These results imply that the LOPEX priors should greatly aid the process of model inversion, by reducing the uncertainty on some of the model inputs, allowing the data to more

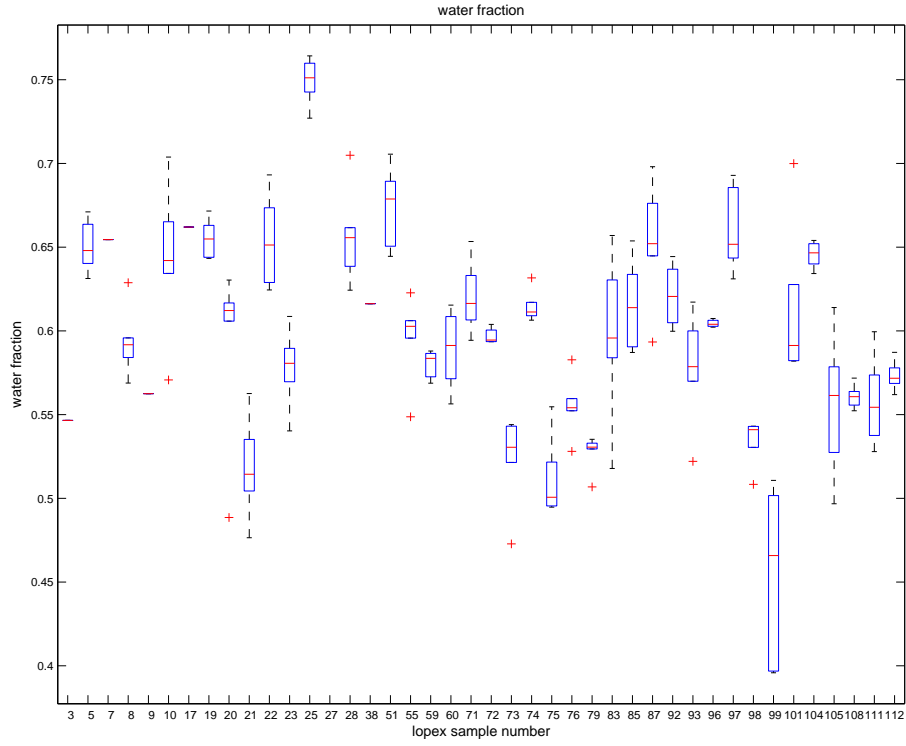


Figure 3: The distributions of the water fractions in each leaf type in the LOPEX database.

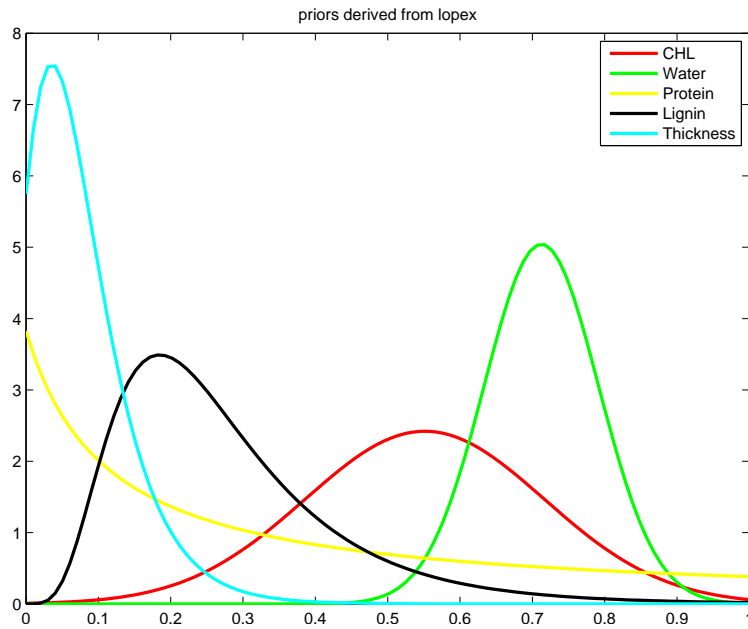


Figure 4: Priors derived from the LOPEX database. Range of each variable is rescaled to 0-1.

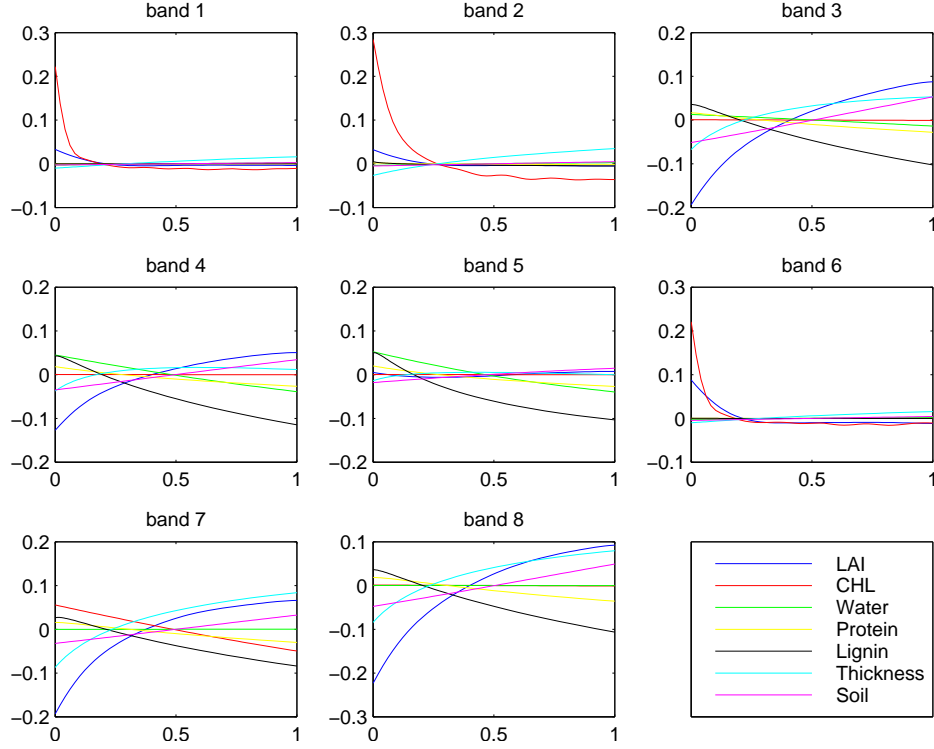


Figure 5: Main Effects for Planophile canopy; uniform priors.

tightly constrain the LAI input.

3 Completion of the Development of the fully-Bayesian Approach to Sensitivity Analysis

As discussed in the IEEE Transactions on Geoscience and Remote Sensing paper that we published last year, the methodology we are developing is based on a Gaussian Process (GP) approximation to the LCM. The results to date have used maximum likelihood to estimate the parameters of the GP. This ignores the uncertainty in these parameters from the finite training data used to estimate them.

To incorporate this uncertainty, we developed a Markov chain Monte Carlo (MCMC) algorithm to sample not only from the distribution of the GP parameters, but also from the prediction of the output of the LCM. Using these samples in a straightforward Monte Carlo estimation framework, we can estimate all of the expectations needed to compute the main effects and the sensitivity indices. Taking into account the uncertainty in the GP parameters results in a full distribution over the main effects and sensitivity indices. Full details of this new statistical methodology are given in the draft paper “Fully Bayesian inference for Variance-Based Sensitivity Analysis of Radiative Transfer Models”, included in the appendix, which will be submitted to Technometrics.

In figure 7 we show box plots representing the *distributions* of the sensitivity indices. In magenta we show the first-order sensitivity indices (considering each variable alone), and in cyan we show

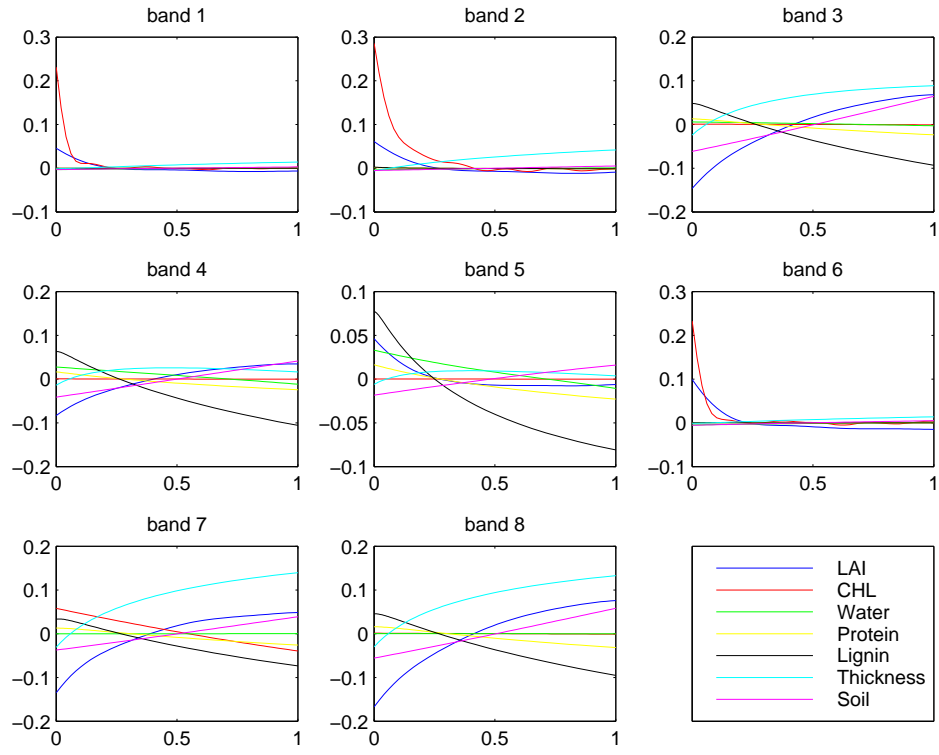


Figure 6: Main Effects for Planophile canopy; LOPEX priors

	band; wavelength (<i>nm</i>)							
input	1 469	2 555	3 1240	4 1640	5 2130	6 667	7 748	8 870
LAI	0.05	0.02	0.50	0.30	0.01	0.23	0.42	0.53
CHL	0.74	0.80	0.00	0.00	0.00	0.59	0.09	0.00
Water	0.00	0.00	0.01	0.09	0.16	0.00	0.00	0.00
Protein	0.00	0.00	0.01	0.02	0.04	0.00	0.02	0.02
Lignin	0.00	0.00	0.14	0.30	0.54	0.00	0.10	0.13
Thickness	0.04	0.07	0.12	0.04	0.01	0.03	0.24	0.17
Soil	0.00	0.00	0.09	0.06	0.03	0.00	0.04	0.07
Total	0.83	0.89	0.88	0.81	0.79	0.85	0.90	0.91

Table 3: Sensitivity Indices for Planophile canopy; uniform priors.

	band; wavelength (<i>nm</i>)							
input	1 469	2 555	3 1240	4 1640	5 2130	6 667	7 748	8 870
LAI	0.38	0.44	0.49	0.29	0.09	0.63	0.45	0.55
CHL	0.39	0.32	0.00	0.00	0.00	0.15	0.05	0.00
Water	0.00	0.00	0.00	0.00	0.01	0.00	0.00	0.00
Protein	0.00	0.00	0.02	0.04	0.09	0.00	0.03	0.03
Lignin	0.00	0.00	0.09	0.24	0.51	0.00	0.07	0.08
Thickness	0.00	0.04	0.06	0.04	0.02	0.00	0.15	0.09
Soil	0.01	0.01	0.20	0.18	0.07	0.01	0.11	0.15
Total	0.78	0.83	0.86	0.80	0.79	0.80	0.86	0.89

Table 4: Sensitivity indices for Planophile canopy; LOPEX priors

the total sensitivity indices – these include the interactions of the stated variable with all the others. We see that in general only some of the variables have significant first-order sensitivity indices in a given band, but that the total sensitivity index can be large, even when the first-order sensitivity index is close to zero. This confirms and quantifies the importance of the interactions, and also the requirement for observations in multiple spectral bands when performing inversion.

4 Completion of a major review of the retrieval of vegetation canopy biophysical parameters from remote sensing data and radiative transfer models

Also in the appendix we include a copy of the paper “Retrieval of Vegetation Canopy Biophysical Parameters via Remote Sensing and Radiative Transfer models: An Overview”, which has been submitted to Measurement Science and Technology as a solicited paper. This provides a major review of the current state of the art in remote sensing model inversion, and a benchmark against which to compare our work.

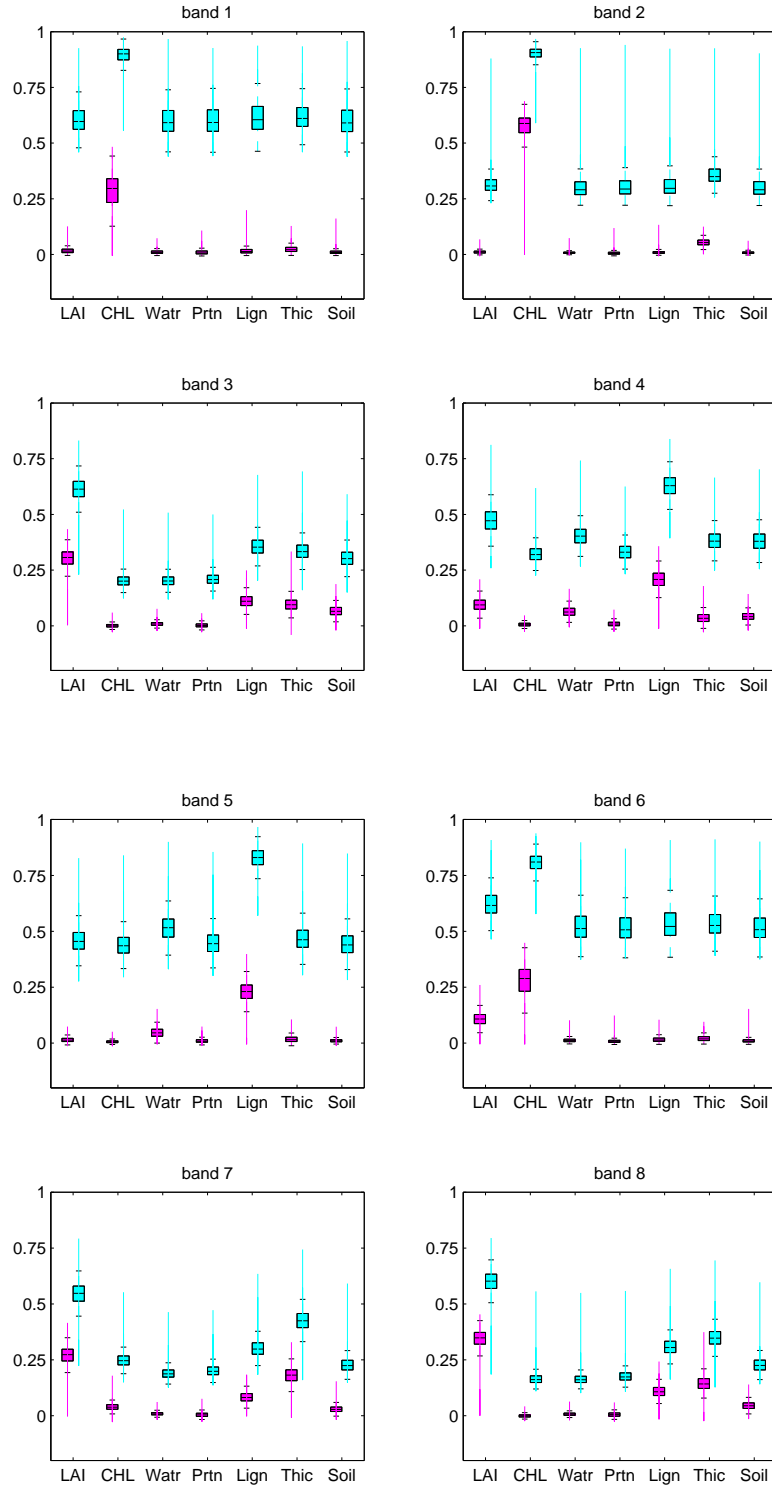


Figure 7: The distributions of the first-order sensitivity indices (magenta) and the total sensitivity indices (cyan) of the LCM inputs as estimated by the fully-Bayesian GP emulator.

5 The award of a Outreach Supplement, which will be used to produce a short video for broadcast on PBS and distribution via YouTube

We submitted a proposal to the ROSES 2008 Outreach Supplemental Awards element (NNH08ZDA001N - OUTREACH), to produce a 5 minute video titled “Living in an Uncertain World”, which will be broadcast on KCSM, one of the local PBS stations in the San Francisco bay area, and potentially on other PBS stations nationwide. It will also be made available on YouTube, in both English and with Spanish subtitles. It will explain, to a lay audience, the main concepts of uncertainty estimation using examples taken from both everyday life (e.g. the lottery) and the science domain of the grant (e.g. uncertainty in climate change forecasts). It will explain the different usage of uncertainty in the two domains, and how accurate uncertainty estimation is vital for accurate decision making.

The supplemental funding was received in early July, and production of the video is scheduled to begin in September.

The technical section of the proposal is attached to this report in the Appendix.

6 Plans for Year Three

Work in year three will address the project’s ultimate goal of model inversion to estimate biospheric parameters from remote sensed data, by completing

- the GP priors for the bias function, which models the discrepancy between the LCM and reality
- the estimation of the bias function using BIGFOOT and MODIS data, providing validation of the model
- inversion of the LCM using the validated GP in the likelihood function.

We will also prepare papers and make presentations as appropriate.

7 Presentations and Publications

- “Extending the Statistical Framework for Global Sensitivity Analysis of Canopy Radiative Transfer Models” A poster presented at the *AGU Fall Meeting*, San Francisco, 2008.
- “A Bayesian Approach to Probabilistic Sensitivity Analysis for Radiative Transfer Models”, Presented at the 2009 SIAM Conference on Computational Science and Engineering, Minisymposium on Bayesian Formulations for System Identification, Modeling and Prediction, Miami, FL, March 2009.
- “Fully Bayesian Inference for Variance-Based Sensitivity Analysis of Radiative Transfer Models”, to be submitted to *Technometrics*.

- “Retrieval of Vegetation Canopy Biophysical Parameters via Remote Sensing and Radiative Transfer models: An Overview”, A solicited paper submitted to Measurement Science and Technology.
- “On the special features of the photon transport in canopies”, 21st International Conference of Transport Theory, Torino, Italy, July 12-17, 2009.
- “Bayesian Estimation and Sensitivity Analysis of Radiative Transfer Models”. A talk at the Mathematical and Computational Biology Seminar Series, UCSC. November, 20, 2008.
- “Bayesian Estimation and Sensitivity Analysis of Radiative Transfer Models”. Poster for Research Review Day, Baskin School of Engineering, UCSC. October 17, 2008.
- “Bayesian Inference for Global Sensitivity Analysis of Radiative Transfer Models”. Poster for the School of Engineering Open House, UCSC. March 12, 2009.

A Appendix: Publications, Posters and Proposals

Copies of the publications and posters listed in section 7 are included in this appendix. The technical section of the supplemental outreach proposal (“Living in an Uncertain World: a video for public television”) is also included here.

Fully Bayesian Inference for Variance-Based Sensitivity Analysis of Radiative Transfer Models

Marian Farah and Athanasios Kottas *

Abstract

Sensitivity analysis is a useful tool for quantifying and describing the sensitivity of a computer model’s output to uncertainty in its inputs. This work is concerned with sensitivity analysis of Radiative Transfer Models (RTMs), which are used for the prediction of a satellite observation of a vegetated region. We consider the Leaf-Canopy Model (LCM), a specific RTM, which takes as input various biospheric and illumination parameters and calculates the upwelling radiation at the top of the canopy. The influence of each input and how uncertainty in the output is apportioned amongst the inputs are determined by calculating the “main effects” and “sensitivity indices” of the LCM inputs. Computing these quantities through Monte Carlo methods using LCM runs is computationally expensive. Using a Bayesian approach, we approximate the LCM by a Gaussian process emulator, and efficiently obtain point estimates of the main effects of the inputs, as well as full posterior distributions of the main effects and sensitivity indices. These distributions are used to determine inputs that are most influential with regard to output prediction uncertainty. We apply these methods to LCM data obtained at 8 wavelengths, which are associated with MODIS (a key instrument aboard the Terra and Aqua satellites) spectral bands that are sensitive to vegetation.

KEY WORDS: Gaussian process prior; Leaf-Canopy Model; Main effects; Sensitivity indices.

*Marian Farah (marian@ams.ucsc.edu) is Ph.D. student, and Athanasios Kottas (thanos@ams.ucsc.edu) is Associate Professor, Department of Applied Mathematics and Statistics, University of California, Santa Cruz, CA 95064, USA. The authors wish to thank Robin Morris, Matthew Taddy, and Roberto Furfaro for helpful discussions. This work was supported in part by the NASA AISR program under grant NNX07AV69G.

1 Introduction

Global process models are widely used in geoscience and remote sensing for the estimation and prediction of the properties of Earth’s coupled dynamical system. Such models are typically implemented in complex computer programs that require global inputs. We are concerned with Radiative Transfer Models (RTMs), which simulate light reflected off the surface of the earth as measured by sensors mounted on orbiting satellites. While RTMs are deterministic models, i.e., they always produce the same output given the same inputs, there is uncertainty about the true values of their inputs. The consequences of this uncertainty in the inputs on the computed output of the RTM is the subject of this paper. We use the Leaf-Canopy Model (LCM) (Ganapol et al., 1999), which simulates light reflected by vegetation, as a surrogate for the RTM, and study the sensitivity of the LCM’s output to uncertainty in its inputs. See Section 2 for a discussion of the LCM.

Sensitivity analysis is an invaluable tool in model development, calibration, and validation. It is concerned with investigating how uncertainty in the model inputs impacts uncertainty in the model output. Thus, sensitivity analysis helps identify where a model must be improved, and where better input information must be obtained. In general, there are two approaches for sensitivity analysis, local and global. Local sensitivity analysis determines how the output changes as the inputs are each varied about a fixed point, while global sensitivity analysis determines how the output changes as all the inputs vary continuously over the entire input space; see Satelli et al. (2000) for a review of both approaches. The sensitivity of the model output is measured via the calculation of the “main effects,” which graphically provide a summary of the influence of each input on the model output, and by the “sensitivity indices,” which are variance-based measures that give the expected amount by which the uncertainty in the output would be reduced if the true value of the input was known (Oakley and O’Hagan, 2004).

Calculating the main effects and sensitivity indices requires the evaluation of multidimensional integrals over the input space of the computer model. Typically, evaluating RTMs is computationally expensive, so consequently, standard numerical integration methods (e.g., Monte Carlo integration or multidimensional quadrature) are infeasible since each model run takes an appreciable amount of time to complete. We adopt the approach of approximating the RTM

by a Gaussian Process (GP), a more computationally efficient model, which can be constructed using a small number of carefully chosen RTM runs; see Section 3. Using the GP approximation (referred to as the GP *emulator*) instead of the actual RTM will introduce uncertainty into the evaluation of the main effects and sensitivity indices, but this uncertainty can be quantified since the GP is a fully specified statistical model. Using GP emulators for computer model experiments is a standard approach in the statistical literature dating back to the work of Sacks et al. (1989); see, e.g., the book by Santner et al. (2003). Moreover, in recent years there has been an upsurge in research activity based on Bayesian methods, see, e.g., Kennedy and O’Hagan (2001), Craig et al. (2001), Oakley and O’Hagan (2002), Higdon et al. (2004), Goldstein and Rougier (2006), Bayarri et al. (2007), Bayarri et al. (2009).

We construct the GP emulator as a prior for the LCM and obtain its posterior distribution; see Section 3. The analysis is fully Bayesian with predictive inference based on the posterior of all model parameters, including the GP prior “smoothness” parameters. In Section 4, we use runs of the GP posterior predictive distribution to obtain full inference for global sensitivity analysis. We first develop an approach to calculate Bayesian point estimates of the main effects and their associated uncertainties. Next, we design a method to obtain full posterior distributions of the main effects and sensitivity indices over the input space of the model. Previous work, such as the semi-Bayesian approach described in Oakley and O’Hagan (2004) and the likelihood approach used by Morris et al. (2008), relied on ratios of point estimates to obtain (ad-hoc) estimates of the sensitivity indices, which are defined through ratios of variances. In Section 5, we present inference for the main effects and sensitivity indices of the LCM and determine the relative importance of each input to the LCM output at 8 different MODIS bands. Finally, Section 6 concludes with discussion of possible extensions.

2 Leaf-Canopy Model

The LCM was developed by the Vegetation Modeling Transport Group (University of Arizona), in collaboration with the Ecosystem Science and Technology Branch at NASA Ames in support of MODIS (or Moderate Resolution Imaging Spectroradiometer), a key instrument aboard Terra and Aqua satellites. The LCM combines two different radiative transfer algo-

rithms: LEAFMOD, which simulates the radiative regime inside the single leaf, and CANMOD, which combines the information coming from LEAFMOD with canopy structural parameters to compute the radiative regime within and at the top of the canopy. LEAFMOD is calibrated over the LOPEX leaf species archive, which stores experimentally obtained data of leaf properties (Osgood et al., 1995). The first module uses LEAFMOD in the forward and inverse mode to compute the leaf optical properties (i.e., leaf reflectance and transmittance). The second module uses the CANMOD forward mode to compute the canopy hemispherical reflectance factor (Ganapol et al., 1999). Figure 1 shows a flowchart that demonstrates the operation of the LCM.

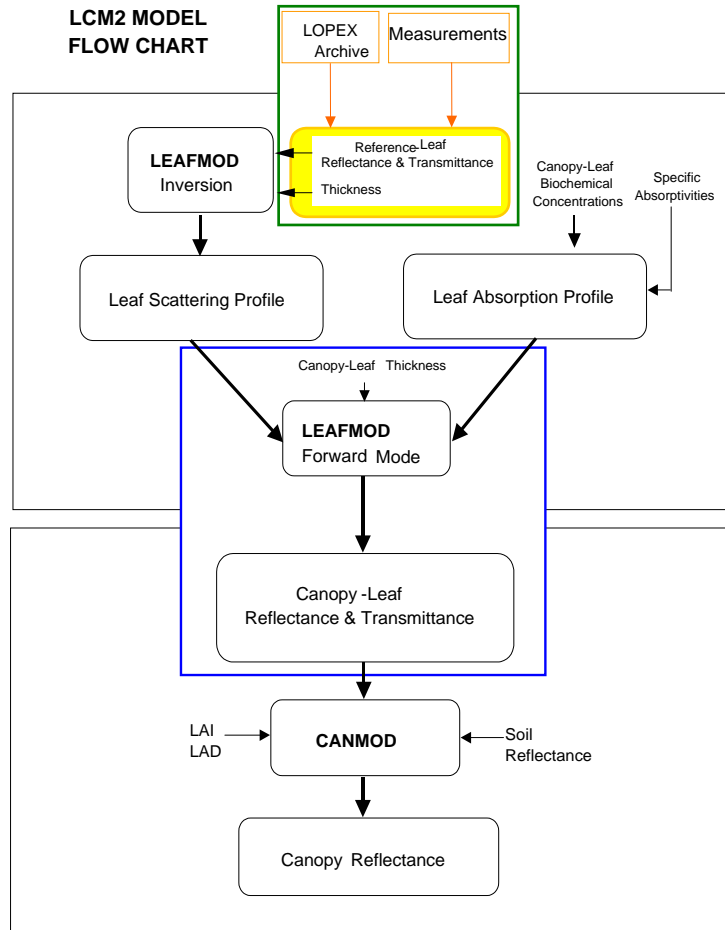


Figure 1: LCM Flow Chart

The LCM requires the specification of the input variables, which include leaf chemistry variables (chlorophyll, water fraction, lignin/cellulose, and protein), leaf thickness, soil reflectance, canopy architecture (Leaf Area Index (LAI) and Leaf Angle Distribution (LAD)), wavelength,

and sun angle (Ganapol et al., 1999). LAI is the area of the leaves on a canopy divided by the area of the ground covered by the canopy, and so it is a dimensionless quantity. LAD describes the orientation of the leaves, and it takes 5 discrete values: planophile (leaves mainly horizontal), erectophile (leaves mainly vertical), plagiophile (leaves mainly at 45 degrees, extremophile (leaves mainly both horizontal and vertical), unophile (leaves are mainly spherical) (Ganapol et al., 1999). In our analysis, the LAD variable is set to planophile, and the sun angle is set to zenith. In future work, we will use a hierarchical extension of our model by including the other LAD variables in the analysis. Table 1 lists the LCM inputs and their ranges, and Table 2 lists the 8 bands (or groups of wavelengths) used by the LCM along with their corresponding MODIS band numbers.

Input	Min	Max
LAI	0	8
Chlorophyll ($\mu g/cm^2$)	0	100
Water fraction	0.1	0.8
Protein (mg/cm^2)	0.1	1
Lignin/Cellulose (mg/cm^2)	0.1	6
Thickness (cm)	0.001	0.01
Soil reflectance	0.3	1.3

Table 1: Ranges of inputs to the LCM. LAI, water fraction, and soil reflectance parameters are dimensionless

band number	wavelength (nm)	MODIS band
1	469	ref3
2	555	ref4
3	1240	ref5
4	1640	ref6
5	2130	ref7
6	667	ref13
7	748	ref15
8	870	ref16

Table 2: Wavelength for each band used and the corresponding MODIS band number. Bands are in the MODIS band order, not in the wavelength order.

Sensitivity analysis of the LCM identifies inputs that are the main contributors to variability in the computed observation. Thus, it provides guidance as to where better input information must be obtain in order to reduce variability in the output. Additionally, identification of im-

portant inputs gives information as to how well these inputs can be predicted from observations of the model output at different wavelengths.

3 Gaussian Process Emulator for the LCM

An emulator is a statistical model that is used to approximate a computer model; see Rasmussen and Williams (2006), Neal (1998), and Mackay (1998). Given a set of training model runs $D = \{(\mathbf{x}_i, y_i = f(\mathbf{x}_i)) : i = 1, \dots, n\}$, the emulator treats a process model (such as LCM) as a black box and uses D to estimate $f(\cdot)$. The advantage of using a GP emulator is that it is a fully specified statistical model that requires one carefully chosen set of model runs, and so it is both tractable and efficient. While this approximation introduces uncertainty into the computation of the main effects and sensitivity indices, this uncertainty is quantifiable.

A GP is a stochastic process over functions. Under a GP prior for function $f(\cdot)$, for any finite set of input points $(\mathbf{v}_1, \dots, \mathbf{v}_N)$, where $\mathbf{v}_i = (v_{1i}, \dots, v_{ki})$ for $i = 1, \dots, N$, the joint probability distribution of the outputs $(f(\mathbf{v}_1), \dots, f(\mathbf{v}_N))$ is multivariate normal. Furthermore, GP models typically assume that the output is a smooth function of its inputs, that is, nearby locations in the input space produce outputs that are stochastically close in value. A GP is fully specified by its mean function, $E(f(\mathbf{v}))$, and positive definite covariance function, $\text{Cov}(f(\mathbf{v}_i), f(\mathbf{v}_j))$. The flexibility of choosing and adapting the mean and covariance functions allows a GP model to be used to approximate a wide spectrum of functions.

In our analysis we assume constant mean function, $\mu(\mathbf{v}) = \mu$, and consider a covariance function, $\text{Cov}(f(\mathbf{v}_i), f(\mathbf{v}_j)) = \tau^2 \text{Corr}(f(\mathbf{v}_i), f(\mathbf{v}_j))$, which is taken to be isotropic with constant variance, τ^2 , and a correlation function that has the product exponential form

$$\text{Corr}(f(\mathbf{v}_i), f(\mathbf{v}_j)) = \exp \left\{ - \sum_{\ell=1}^k \phi_{\ell} |v_{\ell i} - v_{\ell j}| \right\}, \quad (1)$$

where $\boldsymbol{\phi} = (\phi_1, \dots, \phi_k)$, with $\phi_{\ell} > 0$, is the vector of smoothness parameters of the GP.

3.1 Building the GP Emulator

To obtain the set of training data, D , for the GP emulator, we use a Latin Hypercube design (McKay et al., 1979) to generate the design matrix of the seven inputs to the LCM, that is, {LAI, chlorophyll, water, protein, lignin, thickness, soil reflectance}, and calculate the corresponding outputs (hemispherical reflectance) using the LCM.

We treat the functional form of the LCM, say $f(\cdot)$, as unknown and develop a prior for it in the form of a noisy GP. That is, if $\mathbf{x}_i = (x_{1i}, \dots, x_{7i})$ is the i -th observed design point from the input space of the LCM, and y_i is the corresponding LCM output, then we model $y_i = f(\mathbf{x}_i)$ through $y_i = \hat{f}(\mathbf{x}_i) + \varepsilon_i$. Here, the ε_i are i.i.d. $N(0, J)$ error terms, and $\hat{f}(\cdot)$ is assigned an isotropic GP prior with constant mean, μ , constant variance, τ^2 , and the correlation function in (1). While the LCM computer code is deterministic, we choose to treat it as stochastic albeit with a small controlled amount of noise, specified through fixed variance, J . This is a standard computational strategy in Bayesian analysis of computer model experiments with GP priors, and it is analogous to including a *jitter* term in the covariance structure in GP regression to prevent the correlation matrix from being singular (e.g., Neal, 1998).

Given D , there are n induced variables from the GP for $\hat{f}(\cdot)$, denoted by $\theta_i = \hat{f}(\mathbf{x}_i)$, for $i = 1, \dots, n$, with induced prior $\boldsymbol{\theta} = (\theta_1, \dots, \theta_n) \sim N_n(\mu \mathbf{1}_n, \tau^2 R_\phi)$. Here, $\mathbf{1}_n$ is the n -dimensional vector with all elements equal to 1, and R_ϕ is the observed correlation matrix with (i, j) -th element given by $\exp\left\{-\sum_{\ell=1}^7 \phi_\ell |x_{\ell i} - x_{\ell j}|\right\}$, for $i = j = 1, \dots, n$.

We place a normal prior on μ , an inverse-gamma prior on τ^2 , and a $\text{Unif}(0, b_{\phi_\ell})$ prior on each ϕ_ℓ , $\ell = 1, 2, \dots, 7$, assuming prior independence for all hyperparameters. Details on prior specification as well as MCMC posterior simulation for the GP model parameters are provided in Appendix A.

3.2 The Posterior Predictive Distribution of the GP Emulator

Analysis of computer model output performed using runs of the emulator have an additional measure of uncertainty, since those runs are an approximation of the computer code output. We account for this uncertainty by performing further analysis over the posterior predictive distribution of the GP. For any generic input, $\mathbf{v} = (v_1, \dots, v_7)$, we can obtain the posterior

predictive distribution for $\tilde{y} = f(\mathbf{v})$ through the predictive distribution for $\tilde{\theta} = \hat{f}(\mathbf{v})$. Specifically,

$$p(\tilde{y}|D) = \int N(\tilde{y} | \tilde{\theta}, J) p(\tilde{\theta} | D) d\tilde{\theta} = \iint N(\tilde{y} | \tilde{\theta}, J) N(\tilde{\theta} | m(\mathbf{v}), S(\mathbf{v})) p(\boldsymbol{\psi}|D) d\boldsymbol{\psi} d\tilde{\theta}, \quad (2)$$

where $\boldsymbol{\psi} = (\boldsymbol{\theta}, \mu, \tau^2, \boldsymbol{\phi})$, and

$$m(\mathbf{v}) \equiv E(\tilde{\theta} | \boldsymbol{\psi}) = \mu + \mathbf{r}^T(\mathbf{v}) R_{\boldsymbol{\phi}}^{-1} (\boldsymbol{\theta} - \mu \mathbf{1}_n), \quad (3)$$

$$S(\mathbf{v}) \equiv Var(\tilde{\theta} | \boldsymbol{\psi}) = \tau^2 \left(1 - \mathbf{r}^T(\mathbf{v}) R_{\boldsymbol{\phi}}^{-1} \mathbf{r}(\mathbf{v}) \right), \quad (4)$$

where $\mathbf{r}(\mathbf{v})$ is the $n \times 1$ vector with i -th element given by $\exp\left(-\sum_{\ell=1}^7 \phi_{\ell} |v_{\ell} - x_{\ell i}|\right)$.

The joint predictive distribution for $(\tilde{\theta}, \tilde{\theta}') = (\hat{f}(\mathbf{v}), \hat{f}(\mathbf{v}'))$ corresponding to generic inputs $\mathbf{v} = (v_1, v_2, \dots, v_7)$ and $\mathbf{v}' = (v'_1, v'_2, \dots, v'_7)$ is given by $p(\tilde{\theta}, \tilde{\theta}' | D) = \int p(\tilde{\theta}, \tilde{\theta}' | \boldsymbol{\psi}) p(\boldsymbol{\psi} | D) d\boldsymbol{\psi}$, where $p(\tilde{\theta}, \tilde{\theta}' | \boldsymbol{\psi})$ is bivariate normal with (2×1) mean vector

$$\boldsymbol{\omega}(\mathbf{v}, \mathbf{v}') = \mu \mathbf{1}_2 + R^T(\mathbf{v}, \mathbf{v}') R_{\boldsymbol{\phi}}^{-1} (\boldsymbol{\theta} - \mu \mathbf{1}_n), \quad (5)$$

and (2×2) covariance matrix

$$C(\mathbf{v}, \mathbf{v}') = \tau^2 \left(B(\mathbf{v}, \mathbf{v}') - R^T(\mathbf{v}, \mathbf{v}') R_{\boldsymbol{\phi}}^{-1} R(\mathbf{v}, \mathbf{v}') \right), \quad (6)$$

where $B(\mathbf{v}, \mathbf{v}')$ is the (2×2) observed correlation matrix for $(\hat{f}(\mathbf{v}), \hat{f}(\mathbf{v}'))$ with off diagonal elements given by $\exp\left(-\sum_{\ell=1}^7 \phi_{\ell} |v_{\ell} - v'_{\ell}|\right)$, and $R(\mathbf{v}, \mathbf{v}')$ is the $(n \times 2)$ matrix with first column elements $\exp\left(-\sum_{\ell=1}^7 \phi_{\ell} |v_{\ell} - x_{\ell i}|\right)$, $i = 1, 2, \dots, n$, and analogously for the second column elements replacing v_{ℓ} with v'_{ℓ} .

The joint predictive distribution for $(\tilde{y}, \tilde{y}') = (f(\mathbf{v}), f(\mathbf{v}'))$ is given by

$$p(\tilde{y}, \tilde{y}' | D) = \iint p(\tilde{y}, \tilde{y}' | \tilde{\theta}, \tilde{\theta}') p(\tilde{\theta}, \tilde{\theta}' | \boldsymbol{\psi}) p(\boldsymbol{\psi} | D) d\boldsymbol{\psi} d\tilde{\theta} d\tilde{\theta}',$$

where $p(\tilde{y}, \tilde{y}' | \tilde{\theta}, \tilde{\theta}')$ is bivariate normal with mean $(\tilde{\theta}, \tilde{\theta}')$ and covariance matrix $J I_2$.

4 Sensitivity Analysis

4.1 Main Effects and Sensitivity Indices

The key idea behind variance-based approaches to sensitivity analysis is to decompose the function $y = f(\mathbf{v})$ into summands of increasing dimensionality Sobol (1993). Specifically, given a k dimensional input space,

$$y = f(\mathbf{v}) = f_0 + \sum_{i=1}^k f_i(v_i) + \sum_{1 \leq i < j \leq k} f_{i,j}(v_i, v_j) + \cdots + f_{1,2,\dots,k}(v_1, \dots, v_k).$$

Here, f_0 is the global mean given by

$$f_0 = E(Y) = \int_{\mathbf{v}} f(\mathbf{v}) dH(\mathbf{v}),$$

where $H(\mathbf{v}) = \prod_{\ell=1}^k H_{\ell}(v_{\ell})$ is the uncertainty distribution of the inputs comprising independent components $H_{\ell}(v_{\ell})$. For the LCM model, we assume independent uniform distributions over the ranges given in Table 1 for each input variable.

The next k terms are the main effects, where $f_i(v_i)$ is the main effect of input v_i , providing a measure of the influence of input v_i on the computed output. For $i = 1, \dots, k$, we have, in general,

$$f_i(v_i) = E(Y|v_i) - E(Y) = \int_{\mathbf{v}_{-i}} f(\mathbf{v}) dH(\mathbf{v}_{-i}|v_i) - E(Y),$$

where \mathbf{v}_{-i} denotes all elements of \mathbf{v} except v_i . Because of the independent components of the uncertainty distribution, the conditional distribution $H(\mathbf{v}_{-i}|v_i)$ simplifies to $H(\mathbf{v}_{-i})$.

The later terms of the decomposition are the interactions, which quantify the combined influence on the computed output of two or more inputs taken together. For instance, the first-order interactions, $f_{i,j}(v_i, v_j) = E(Y|v_i, v_j) - f_i(v_i) - f_j(v_j) - E(Y)$.

Sobol (1993) shows that based on this output decomposition, and assuming independence between the input variables in the uncertainty distribution, the total variance, $\text{Var}(Y) = W$,

can also be decomposed as the sum of partial variances,

$$W = \sum_{i=1}^k W_i + \sum_{1 \leq i < j \leq k} W_{i,j} + \cdots + W_{1,2,\dots,k}, \quad (7)$$

where $W_i = \text{Var}(f_i(v_i)) = \text{Var}\{E(Y|v_i)\}$, $W_{i,j} = \text{Var}(f_{i,j}(v_i, v_j)) = \text{Var}\{E(Y|v_i, v_j)\} - W_i - W_j$, and analogously for the higher order terms. Hence, the sensitivity indices, are given by

$$S_i = \frac{W_i}{W}, \quad S_{i,j} = \frac{W_{i,j}}{W}, \quad \dots, \quad S_{1,2,\dots,k} = \frac{W_{1,2,\dots,k}}{W},$$

where S_i is called the *first-order sensitivity index* for input v_i , which measures the fractional contribution of v_i to the variance of $f(\mathbf{v})$, $S_{i,j}$, for $i \neq j$, is called the *second-order sensitivity index*, which measures the contribution of interaction due to v_i and v_j on the variance of $f(\mathbf{v})$, and so on. The decomposition in (7) standardizes the sensitivity indices, that is,

$$\sum_{i=1}^k S_i + \sum_{1 \leq i < j \leq k} S_{i,j} + \cdots + S_{1,2,\dots,k} = 1$$

Introduced by Homma and Satelli (1996), the *total sensitivity index*, S_i^T , is a further related measure, defined by the sum of all the sensitivity indices involving input v_i . Specifically,

$$S_i^T = 1 - \frac{W_{-i}}{W}, \quad i = 1, \dots, k,$$

where $W_{-i} = \text{Var}\{E(Y|\mathbf{v}_{-i})\}$ is the total contribution to the variance of $f(\mathbf{v})$ due to all inputs except v_i .

Note that since we are approximating the function $y = f(\mathbf{v})$ by a GP model, we must account for this approximation by computing $E^*\{E(Y)|D\}$, $E^*\{E(Y|v_i)|D\}$, $E^*\{S_j|D\}$, and $E^*\{S_j^T|D\}$, where $E^*\{\cdot|D\}$ indicates the expectations with respect to the GP posterior predictive distribution, $p(\tilde{y}|D)$, developed in Section 3.2.

4.2 Bayesian Point Estimates and Uncertainty Bands for The Main Effects

Let $\mathbf{v} = (v_1, \dots, v_7)$ be generic input of the GP emulator. Then, the distribution of the predicted emulator output, $\tilde{y} = f(\mathbf{v})$, is given by (2).

Assuming independent components in the uncertainty distribution for the inputs, the global mean is given by

$$E(Y) = \int_{\mathbf{v}} f(\mathbf{v}) \prod_{\ell=1}^7 dH_{\ell}(v_{\ell}).$$

Using (2) and (3), we obtain

$$E^* \{E(Y) | D\} = \int_{f(\mathbf{v})} E(Y) p(f(\mathbf{v}) | D) df(\mathbf{v}) = \int_{\boldsymbol{\psi}} \left\{ \mu + \mathbf{T}^T R_{\phi}^{-1} (\boldsymbol{\theta} - \mu \mathbf{1}_n) \right\} p(\boldsymbol{\psi} | D) d\boldsymbol{\psi},$$

where \mathbf{T} is the $n \times 1$ vector with i -th element given by

$$\prod_{\ell=1}^7 \left\{ \int_0^1 \exp(-\phi_{\ell} |v_{\ell} - x_{\ell i}|) dv_{\ell} \right\} = \prod_{\ell=1}^7 \left\{ \frac{2 - e^{-\phi_{\ell} x_{\ell i}} - e^{-\phi_{\ell} (1 - x_{\ell i})}}{\phi_{\ell}} \right\}.$$

For each specified value u_j of the j -th input, we have

$$E(Y|u_j) = \int_{\{v_{\ell}: \ell \neq j\}} f(v_1, \dots, u_j, \dots, v_7) \prod_{\{\ell: \ell \neq j\}} dH_{\ell}(v_{\ell})$$

Again, using (2) and (3), we obtain

$$\begin{aligned} E^* \{E(Y|u_j) | D\} &= \int_{f(v_1, \dots, u_j, \dots, v_7)} E(y|u_j) p(f(v_1, \dots, u_j, \dots, v_7) | D) df(v_1, \dots, u_j, \dots, v_7) \\ &= \int_{\boldsymbol{\psi}} \left\{ \mu + T_j^T(u_j) R_{\phi}^{-1} (\boldsymbol{\theta} - \mu \mathbf{1}_n) \right\} p(\boldsymbol{\psi} | D) d\boldsymbol{\psi}, \end{aligned}$$

where $T_j(u_j)$ is the $n \times 1$ vector with i -th element given by

$$\exp(-\phi_j |u_j - x_{ji}|) \times \prod_{\ell \neq j} \left\{ \int_0^1 \exp(-\phi_{\ell} |v_{\ell} - x_{\ell i}|) dv_{\ell} \right\}.$$

For a measure of posterior predictive uncertainty associated with the estimate of the main effects, we use

$$Var^* \{E(Y|u_j) | D\} = E^* \left\{ (E(Y|u_j))^2 | D \right\} - (E^* \{E(Y|u_j) | D\})^2$$

Because we already have the expression for $E^* \{E(Y|u_j) | D\}$, what is needed is an expression for $E^* \{(E(Y|u_j))^2 | D\}$. Let $\mathbf{v}_j \equiv (v_1, \dots, u_j, \dots, v_7)$ and $\mathbf{v}'_j \equiv (v'_1, \dots, u_j, \dots, v'_7)$. Then, extending the arguments in the derivation of $E^* \{E(Y) | D\}$ and $E^* \{E(Y|u_j) | D\}$, we obtain

$$\begin{aligned} E^* \{(E(Y|u_j))^2 | D\} &= \int (E(Y|u_j))^2 p(f(\mathbf{v}_j), f(\mathbf{v}'_j) | D) df(\mathbf{v}_j) df(\mathbf{v}'_j) \\ &= \int_{\psi} \left\{ \tau^2 \left(e - \mathbf{T}_j^T(u_j) R_{\phi}^{-1} T_j(u_j) \right) + \left(\mu + \mathbf{T}_j^T(u_j) R_{\phi}^{-1} (\boldsymbol{\theta} - \mu \mathbf{1}_n) \right)^2 \right\} p(\psi | D) d\psi, \end{aligned}$$

where

$$e = \prod_{\{\ell: \ell \neq j\}} \left\{ \int_0^1 \int_0^1 \exp(-\phi_{\ell} |v_{\ell} - v'_{\ell}|) dv_{\ell} dv'_{\ell} \right\} = \prod_{\{\ell: \ell \neq j\}} \frac{2}{\phi_j^2} (e^{-\phi_j} + \phi_j - 1)$$

For detailed derivations of $E^* \{E(Y) | D\}$, $E^* \{E(Y|u_j) | D\}$, and $E^* \{(E(Y|u_j))^2 | D\}$, see Appendix B. It should be noted that Morris et al. (2008) use the expressions we have in the integrands of all three derived expressions as maximum likelihood estimates by plugging in maximum likelihood estimators of the GP parameters. The difference here is that we use a Bayesian GP, which takes into account the uncertainty distributions of its parameters. Also, Oakley and O'Hagan (2004) derive similar expressions for the main effects and their associated uncertainty using a GP emulator. However, they use point estimates of the smoothness parameters, while we obtain full posterior distributions for them.

4.3 A Fully-Bayesian Approach to Inference for the Main Effects and Sensitivity Indices

We obtain posterior distributions for the main effects and sensitivity indices by computing the values of $E(Y)$, $\text{Var}(Y)$, $E(Y|u_j)$, and $E((E(Y|u_j))^2)$ at every MCMC state of the GP emulator using Monte Carlo methods reviewed in Satelli (2002). Here, we briefly discuss how we calculate these expressions. Letting $\{\mathbf{v} = (v_1, \dots, v_7), y = f(\mathbf{v})\}$ be a generic run of the LCM,

the expectation and variance of y are given by

$$\begin{aligned}
E(Y) &= \iiint \cdots \int f(v_1, v_2, \dots, v_7) \prod_{\ell=1}^7 p_\ell(v_\ell) dv_\ell. \\
\text{Var}(Y) &= \iiint \cdots \int (f(v_1, v_2, \dots, v_7) - E(Y))^2 \prod_{\ell=1}^7 p_\ell(v_\ell) dv_\ell \\
&= \iiint \cdots \int (f^2(v_1, v_2, \dots, v_7)) \prod_{\ell=1}^7 p_\ell(v_\ell) dv_\ell - (E(Y))^2.
\end{aligned}$$

For a generic value u_j of the j -th input, we have

$$E(Y|u_j) = \iiint \cdots \int f(v_1, \dots, u_j, \dots, v_7) \prod_{\substack{\ell=1 \\ \ell \neq j}}^7 p_\ell(v_\ell) dv_\ell.$$

Squaring this expression, then taking its expectation, we obtain

$$\begin{aligned}
E((E(Y|u_j))^2) &= \int \left\{ \iiint \cdots \int f(v_1, \dots, u_j, \dots, v_7) \prod_{\substack{\ell=1 \\ \ell \neq j}}^7 p_\ell(v_\ell) dv_\ell \right\}^2 p_j(u_j) du_j \\
&= \iiint \cdots \int f(v_1, \dots, u_j, \dots, v_7) f(v'_1, \dots, u_j, \dots, v'_7) \\
&\quad \times \prod_{\ell=1}^7 p_\ell(v_\ell) dv_\ell \prod_{\substack{\ell=1 \\ \ell \neq j}}^7 p_\ell(v'_\ell) dv'_\ell.
\end{aligned}$$

Letting $\mathbf{u}_{-j} = (u_1, \dots, u_{j-1}, u_{j+1}, \dots, u_7)$. Then,

$$E(Y|\mathbf{u}_{-j}) = \int f(v_j, \mathbf{u}_{-j}) p_j(v_j) dv_j,$$

and

$$\begin{aligned}
E \left((E(Y|\mathbf{u}_{-j}))^2 \right) &= \int \cdots \int \left\{ \int f(v_j, \mathbf{u}_{-j}) p_j(v_j) dv_j \right\}^2 \prod_{\substack{\ell=1 \\ \ell \neq j}}^7 p_\ell(u_\ell) du_\ell \\
&= \iint \cdots \int f(v_j, \mathbf{u}_{-j}) f(v'_j, \mathbf{u}_{-j}) p_j(v_j) dv_j p_j(v'_j) dv'_j \prod_{\substack{\ell=1 \\ \ell \neq j}}^7 p_\ell(u_\ell) du_\ell \\
&= \iint \cdots \int f(v_j, \mathbf{u}_{-j}) f(v'_j, \mathbf{u}_{-j}) p_j(v'_j) dv'_j \prod_{\ell=1}^7 p_\ell(u_\ell) du_\ell.
\end{aligned}$$

Thus, at each MCMC state, t , of the GP emulator, the main effect, f_j , the first-order sensitivity index, S_j , and the total sensitivity index, S_j^T are calculated as follows:

$$f_j^{(t)} = E(Y|u_j)^{(t)} - E(Y)^{(t)} \quad (8)$$

$$S_j^{(t)} = \frac{\text{Var}(E(Y|u_j))^{(t)}}{\text{Var}(Y)^{(t)}} = \frac{E(E^2(Y|u_j))^{(t)} - E^2(Y)^{(t)}}{\text{Var}(Y)^{(t)}} \quad (9)$$

$$S_j^{T(t)} = \frac{\text{Var}(Y)^{(t)} - \text{Var}(E(Y|\mathbf{u}_{-j}))^{(t)}}{\text{Var}(Y)^{(t)}} = 1 - \frac{E(E^2(Y|\mathbf{u}_{-j}))^{(t)} - E^2(Y)^{(t)}}{\text{Var}(Y)^{(t)}} \quad (10)$$

While a fresh set of data is required for computing (8), (9), and (10), these are *cheap* emulator runs, which are enormously faster than those obtained using the LCM code. See Appendix C for technical implementation.

Thus, we are able to estimate the entire distributions of the the main effects and sensitivity indices of the LCM inputs under the GP approximation, allowing for the uncertainty in the main effects and sensitivity indices to be accurately quantified. In previous work, Oakley and O'Hagan (2004) approximated $E^*(S_j)$ using the ratio of $E^*\{\text{Var}(E(Y|u_j))\}$ and $E^*\{\text{Var}(Y)\}$. Also, Morris et al. (2008) estimated S_j using a similar ratio based on expectations of the variances that define S_j with respect to the GP predictive distribution, with GP parameters estimated by their maximum likelihood estimates.

5 Results

We apply the Bayesian approach to the GP emulator (Section 3) using a training set of 250 model runs based on a Latin Hypercube design at each band (see Table 2). The LAD

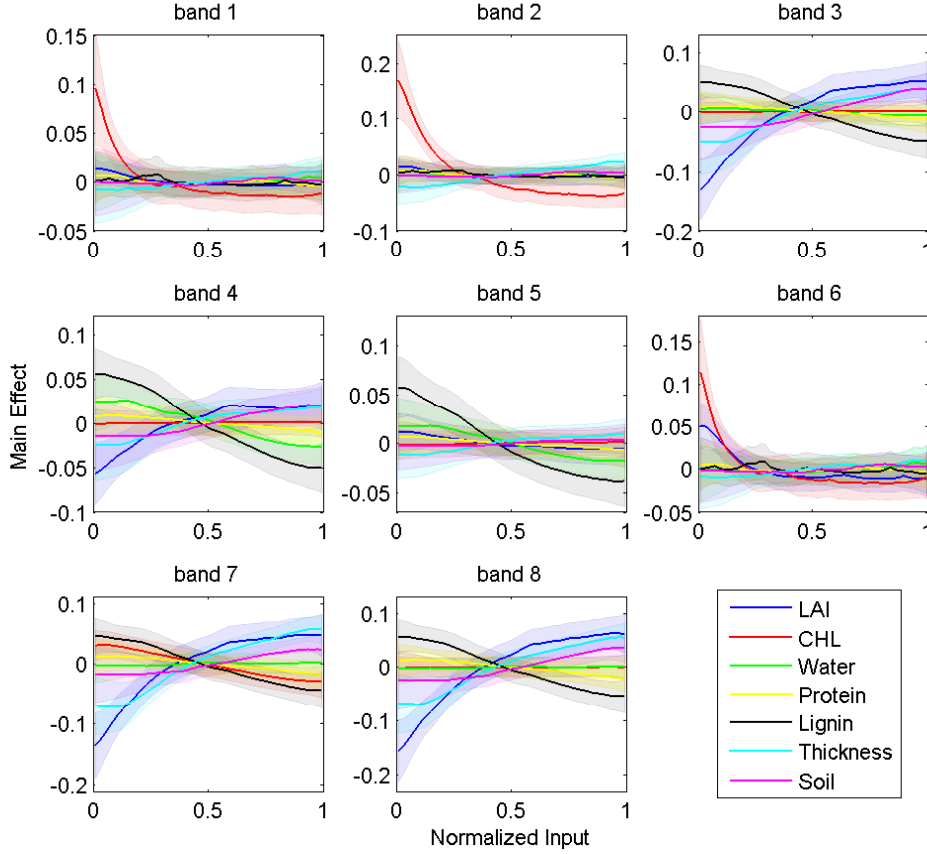


Figure 2: Bayesian point estimates ± 2 standard deviations for the main effects. The heavy smooth lines indicate the point estimates, and the shaded regions around them indicate the associated ± 2 standard deviations region.

variable is set to planophile (leaves mostly horizontal), and the sun angle is set to zenith. Based on posterior GP emulator runs, we obtain point estimates and uncertainty bands of the main effects as well as full posterior distributions of the main effects and sensitivity indices of the LCM inputs, as described in Section 4.

Figure 2 shows plots of the main effects for the 7 normalized input variables and their uncertainty intervals for each of the 8 MODIS bands. Normalizing the inputs allows all the main effects to be plotted together on the same plot. The larger the variation of the main effect plot, the greater the influence of that input on the LCM output. The slope of each main effect plot gives information as to whether the output is an increasing or decreasing function of that input.

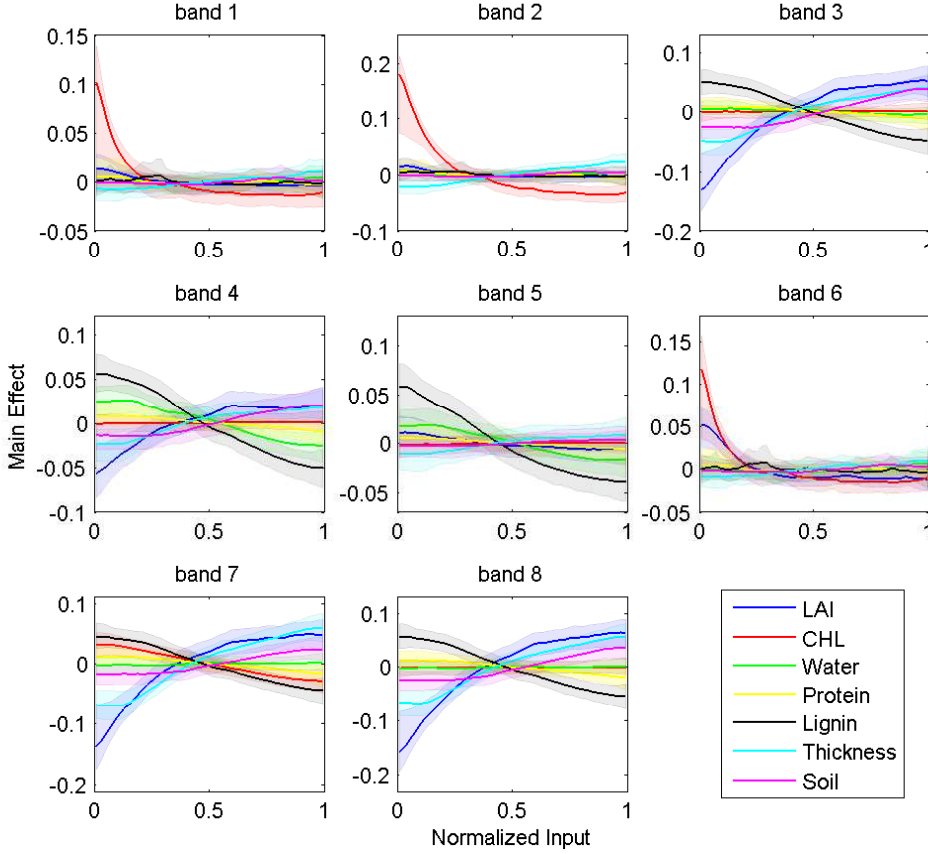


Figure 3: Medians and 95% probability bands of the posterior distributions of main effects. The heavy smooth lines indicate the medians, and the shaded regions around them indicate the associated 95% probability bands.

For visible spectrum (bands 1, 2, & 6), the LCM is most sensitive to chlorophyll, with LAI becoming important for red light (band 6), and an increase in Chlorophyll produces a decrease in the LCM output. For near infra-red (bands 3, 7, & 8), the LCM is most sensitive to LAI, lignin, and thickness, and an increase in LAI or thickness produces an increase in the LCM output, while an increase in lignin produces a decrease in the LCM output. Finally, for short infra-rad bands (bands 4 & 5), the chlorophyll effect is completely diminished, while LAI and lignin continue to dominate, and water becoming more influential. In general, we observe that all dominant inputs have nonlinear main effects. These results are consistent with the results obtained by Morris et al. (2008). However, our estimated uncertainty bands are wider than the confidence intervals obtained by Morris et al. (2008), which can be explained by the fact that

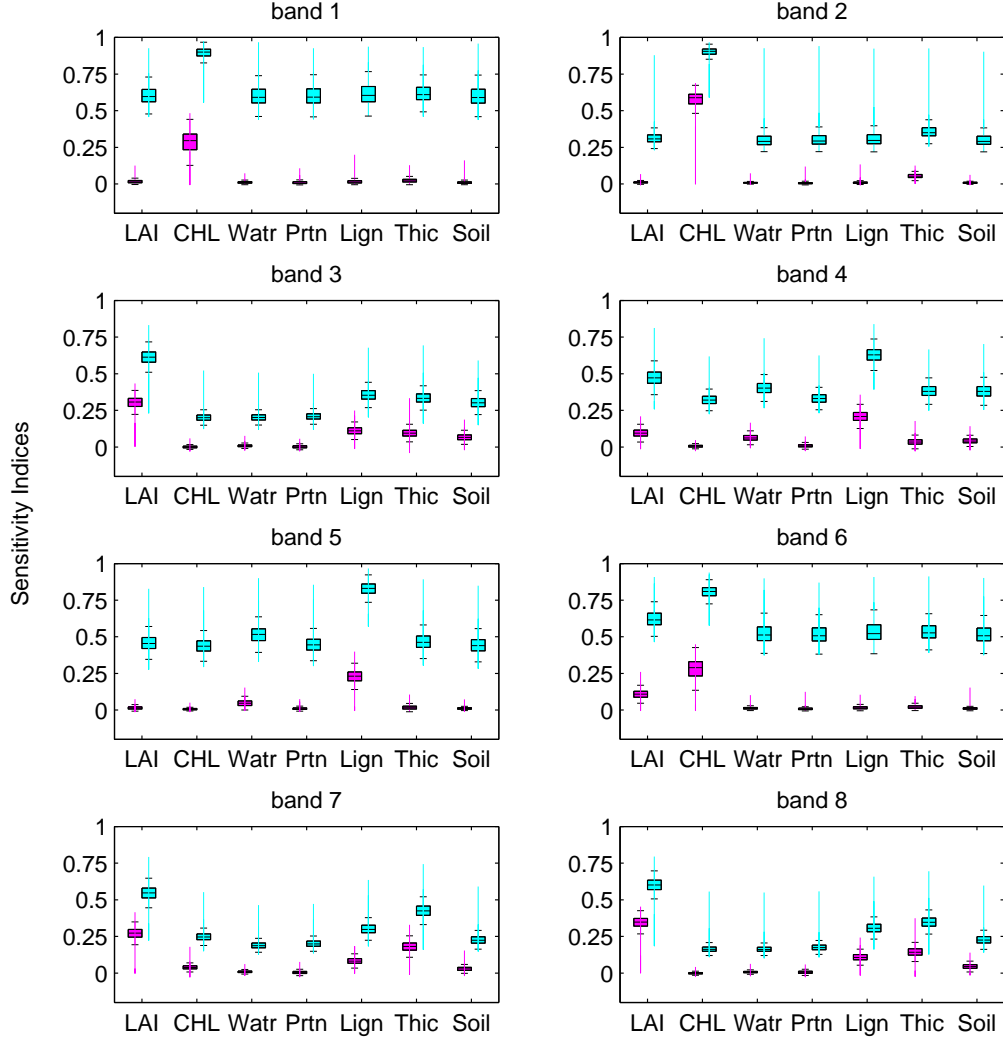


Figure 4: The distributions of the first-order sensitivity indices (in magenta) and the total sensitivity indices (in cyan) of the LCM inputs as estimated by the GP emulator.

we are also accounting for uncertainty in the GP parameters.

Figure 3 shows plots of the main effects, obtained using the fully-Bayesian approach, where posterior distributions instead of point estimates are obtained for the 7 normalized input variables at each of the 8 bands. The means of the distributions of the main effects agree with the point estimates in figure 1, and the influential inputs obtained using point estimates are the same as those obtained through the fully-Bayesian approach. However, we observe that the uncertainty bands become less symmetric around the means in the fully-Bayesian approach for dominant inputs near the boundary of the grid.

Figure 4 shows boxplots of the first-order sensitivity indices and the total sensitivity indices. These plots show that inputs with influential main effects are also major contributors to the variation in the LCM (i.e., they have large sensitivity indices). Additionally, we observe that many inputs with negligible (nearly zero) first order sensitivity indices had non-negligible total sensitivity indices. Specifically, for all bands, the total effect index of every input has over 95% probability of being greater than 0.10. A substantial difference between S_j and S_j^T of the j -th input indicates an important role of interaction terms involving that input on the variation in the output. Thus, we find that interaction terms involving all LCM inputs are influential at all 8 MODIS bands, and so the dimension of the LCM input is irreducible.

6 Conclusion and Future Work

We have implemented a Bayesian approach, via MCMC methods for the GP emulator, to obtain point estimates for all the main effects as well as posterior distributions of the main effects and sensitivity indices associated with the 7 LCM inputs at 8 different MODIS wavelengths. Our analysis enabled the identification of influential first-order effects of the inputs to the LCM and revealed that interaction terms are also important in controlling the variation of the LCM output. We plan to explore some of the second-order sensitivity indices for inputs that had important first-order sensitivity indices.

Additionally, it will be of interest to study the variable selection approach from Linkletter et al. (2006) in the context of sensitivity analysis. Linkletter et al. (2006) use Bayesian variable

selection for the GP correlation parameters to make screening decisions in order to reduce the input space by identifying “active” inputs. Moreover, a different direction for extending the work for the LCM model involves development of a more general GP emulator to account for all 4 LAD classifications, using a hierarchical GP formulation for the corresponding output functions $f_j(v)$, for $j = 1, 2, 3, 4$.

References

- Bayarri, M. J., Berger, J. O., Kennedy, M., Kottas, A., Paulo, R., Sacks, J., Cafeo, J. A., Lin, C. H., and Tu, J. (2009), “Predicting Vehicle Crashworthiness: Validation of Computer Models for Functional and Hierarchical Data,” *Journal of the American Statistical Association* (to appear).
- Bayarri, M. J., Berger, J. O., Paulo, R., Sacks, J., Cafeo, J. A., Cavendish, J., Lin, C. H., and Tu, J. (2007), “A framework for validation of computer models,” *Technometrics*, 49, 138–154.
- Craig, P. S., Goldstein, M., Rougier, J. C., and Seheult, A. H. (2001), “Bayesian forecasting for complex systems using computer simulators,” *Journal of the American Statistical Association*, 96, 717–729.
- Ganapol, B. D., Johnson, L. F., Hlavka, C. A., D.L., P., and Bond, B. (1999), “LCM2: A Coupled Leaf/Canopy Radiative Transfer Model,” *Remote Sensing of the Environment*, 153–166.
- Goldstein, M. and Rougier, J. C. (2006), “Bayes linear calibrated prediction for complex systems,” *Journal of the American Statistical Association*, 101, 1132–1143.
- Higdon, D., Kennedy, M. C., Cavendish, J., Cafeo, J., and Ryne, R. D. (2004), “Combining field data and computer simulations for calibration and prediction,” *SIAM Journal on Scientific Computing*, 26, 448–466.
- Homma, T. and Satelli, A. (1996), “Importance Measures in Global Sensitivity Analysis of Nonlinear Models,” *Reliability Engineering and System Safety*, 52, 1–17.
- Kennedy, M. C. and O’Hagan, A. (2001), “Bayesian calibration of computer models (with discussion),” *Journal of the Royal Statistical Society B*, 63, 425–464.

- Linkletter, C., Bingham, D., Hengartner, N., Higdon, D., and Ye, K. Q. (2006), “Variable Selection for Gaussian Process Models in Computer Experiments,” *Technometrics*, 48, 478–490.
- Mackay, D. (1998), “Introduction to Gaussian Process,” *NATO ASI Series. Neural Networks and Machine Learning*, 168, 133–166.
- McKay, M. D., Beckman, R. J., and Conover, W. J. (1979), “A Comparison of Three Methods for Selecting Values of Input Variables in the Analysis of Output from a Computer Code,” *Technometrics*, 21, 239–245.
- Morris, R. D., Kottas, A., Taddy, M., Furfaro, R., and Ganapol, B. D. (2008), “A Statistical Framework for the Sensitivity Analysis of Radiative Transfer Models,” *IEEE on Geoscience and Remote Sensing*, 46, 4062–4074.
- Neal, R. M. (1998), “Regression and Classification Using Gaussian Process Priors,” *Bayesian Statistics*, 475–501, Bernardo, J., Berger, J., Dawid, A., and Smith, A., eds.
- Oakley, J. and O’Hagan, A. (2002), “Bayesian inference for the uncertainty distribution of computer model outputs,” *Biometrika*, 89, 769–784.
- Oakley, J. E. and O’Hagan, A. (2004), “Probabilistic Sensitivity Analysis of Complex Models: a Bayesian Approach,” *Journal of the Royal Statistical Society, Series B*, 66, 751–769.
- Rasmussen, C. E. and Williams, C. K. I. (2006), *Gaussian Processes for Machine Learning*, MIT Press.
- Sacks, J., Welch, W. J., Mitchell, T. J., and Wynn, H. P. (1989), “Design and analysis of computer experiments (C/R: p423–435),” *Statistical Science*, 4, 409–423.
- Santner, T., Williams, B., and Notz, W. (2003), *The Design and Analysis of Computer Experiments*, Springer-Verlag.
- Satelli, A. (2002), “Making Best Use of Model Evaluations to Compute Sensitivity Indices,” *Computer Physics Communications*, 45, 280–297.
- Satelli, A., Chan, k., and Scott, E. M., e. (2000), *Sensitivity Analysis*, John Wiley and Sons.

Sobol, I. M. (1993), “Sensitivity Estimates For Non Linear Mathematical Models,” *Mathematical Modelling and Computational Experiments*, 1, 407–414.

APPENDIX A

Posterior inference for the GP emulator

We assume prior independence of all hyperparameters and place a $N(a_\mu, b_\mu)$ prior on μ , an $\Gamma^{-1}(a_\tau, b_\tau)$ prior on τ^2 , and $\text{Unif}(0, b_{\phi_\ell})$ priors on ϕ_ℓ , $\ell = 1, 2, \dots, 7$, where b_μ , a_τ , b_τ , and $b_{\phi_\ell} \in \mathbb{R}^+$, and $a_\mu \in \mathbb{R}$. Here, $\Gamma^{-1}(a, b)$ denotes the inverse-gamma distribution with mean $\frac{b}{a-1}$, provided $a > 1$.

Let $\boldsymbol{\phi} = (\phi_1, \dots, \phi_7)$, and $R_\phi = \text{Corr}(f(\mathbf{x}), f(\mathbf{x}')) = \exp \left\{ -\sum_{\ell=1}^7 \phi_\ell |x_\ell - x'_\ell| \right\}$.

Then, the joint posterior distribution of all parameters is:

$$p(\boldsymbol{\theta}, \mu, \tau^2, \boldsymbol{\phi} | D) \propto \left(\prod_{i=1}^n N(y_i | \theta_i, J) \right) \times N_n(\boldsymbol{\theta} | \mu \mathbf{1}_n, \tau^2 R_\phi) p(\mu) p(\tau^2) p(\boldsymbol{\phi})$$

This form results in complete conditional posterior distributions that are easy to sample from for all parameters other than the ϕ_ℓ . In particular, $\boldsymbol{\theta} | \mu, \tau^2, \boldsymbol{\phi}, D \sim N_n(\mathbf{M}_n, \Sigma_n)$ where $\Sigma_n = \left(\frac{I_n}{J} + \frac{R_\phi^{-1}}{\tau^2} \right)^{-1}$ and $\mathbf{M}_n = \Sigma_n \left(\frac{\mathbf{y}}{J} + \frac{\mu R_\phi^{-1} \mathbf{1}_n}{\tau^2} \right)$. Moreover, the posterior full conditional for μ is normal with variance $s = \left(\frac{\mathbf{1}_n^T R_\phi^{-1} \mathbf{1}_n}{\tau^2} + \frac{1}{b_\mu} \right)^{-1}$, and mean $s \left(\frac{\mathbf{1}_n^T R_\phi^{-1} \boldsymbol{\theta}}{\tau^2} + \frac{a_\mu}{b_\mu} \right)$. For τ^2 , the posterior full conditional is $\Gamma^{-1}(A, B)$, with $A = a_\tau + \frac{n}{2}$ and $B = \frac{1}{2} (\boldsymbol{\theta} - \mu \mathbf{1}_n)^T R_\phi^{-1} (\boldsymbol{\theta} - \mu \mathbf{1}_n) + b_\tau$.

Finally, for each ϕ_ℓ , $\ell = 1, \dots, 7$,

$$p(\phi_\ell | \boldsymbol{\theta}, \mu, \tau^2, \boldsymbol{\phi}_{-\ell}, D) \propto |R_\phi|^{-\frac{1}{2}} \exp \left(-\frac{1}{2\tau^2} (\boldsymbol{\theta} - \mu \mathbf{1}_n)^T R_\phi^{-1} (\boldsymbol{\theta} - \mu \mathbf{1}_n) \right) \times 1_{(0, b_{\phi_\ell})}(\phi_\ell)$$

In the MH step for each ϕ_ℓ , we draw ϕ_ℓ^* from a right-truncated exponential proposal distribution with pdf

$$q(\phi_\ell^*) = \frac{d_\ell \exp(-d_\ell \phi_\ell)}{1 - \exp(-d_\ell b_{\phi_\ell})}, \quad (11)$$

so the rate parameter d_ℓ is the only tuning parameter. One way to choose the parameter d_ℓ in the proposal distribution is to let $\tilde{\phi}_\ell$ be an estimate (the MLE or another rough estimate) of ϕ_ℓ . Set $\tilde{\phi}_\ell$ equal to the median of (11), and solve for d_ℓ . In this case, d_ℓ is the solution to $1 + e^{-d_\ell b_{\phi_\ell}} - 2e^{-d_\ell \phi_\ell} = 0$.

Prior Specification

We set $a_\tau = 2$, a value that yields infinite variance for the corresponding inverse-gamma prior (i.e. a relatively noninformative specification).

To specify the hyperparameters b_μ and b_τ , note that for each i ,

$$E(Y_i) = E(E(E(Y_i|\theta_i)|\mu)) = E(E(\theta_i|\mu)) = E(\mu) = a_\mu$$

$$Var(Y_i) = E(Var(Y_i|\theta_i, J)) + Var(E(Y_i|\theta_i, J)) = J + b_\tau + b_\mu \approx b_\tau + b_\mu$$

Now, assume we have a prior guess for the center, c_y , and range, r_y , of the response values. Availability of such information seems realistic in our application. Then, we set

$$a_\mu = c_y \quad \text{and} \quad b_\tau = b_\mu \approx (r_y/4)^2$$

using $2(r_y/4)^2 \approx b_\tau + b_\mu$, with the extra inflation factor 2, and splitting the variance estimate equally between b_τ and b_μ .

Specifying prior information for ϕ_ℓ , $\ell = 1, 2, \dots, 7$ is more difficult. One way to specify $b_{\phi_\ell} = 2E(\phi_\ell)$ is based on the interpretation of ϕ_ℓ : for any fixed α_ℓ , it controls how fast the correlation decays with distance. In particular, for $\alpha_\ell = 1$, $3/\phi_\ell$ is the “range of dependence”, i.e. the value of the distance $d = |x - x'|$ that yields $Corr(y, y') = 0.05$. Hence, we could use, say, $\frac{1}{10} d_{max}$, where $d_{max} = \max|x_\ell - x'_\ell|$, as a rough guess at $3/\phi_\ell$ and specify b_{ϕ_ℓ} from $\frac{1}{10} d_{max} = \frac{3}{b_{\phi_\ell}}$. The factor $\frac{1}{10}$ is arbitrary to some extent. In practice, numerical problems can arise in the MCMC implementation (in matrix inversions) for small generated values of ϕ_ℓ . If that turns out to be the case, we can increase b_{ϕ_ℓ} (and hence decrease the prior probability for small values of ϕ_ℓ) by using a smaller factor.

APPENDIX B

Here, we provide the details for derivations of expressions for $E^* \{E(Y) | D\}$, $E^* \{E(Y|u_j) | D\}$, and $E^* \{(E(Y|u_j))^2 | D\}$.

$$\begin{aligned}
E^* \{E(Y) | D\} &= \int_{f(\mathbf{v})} E(Y) p(f(\mathbf{v}) | D) df(\mathbf{v}) \\
&= \int_{f(\mathbf{v})} E(Y) \left\{ \int_{\boldsymbol{\psi}} \int_{\tilde{\theta}} N(f(\mathbf{v}) | \tilde{\theta}, J) N(\tilde{\theta} | m(\mathbf{v}), S(\mathbf{v})) p(\boldsymbol{\psi} | D) d\tilde{\theta} d\boldsymbol{\psi} \right\} df(\mathbf{v}) \\
&= \int_{\boldsymbol{\psi}} \int_{\tilde{\theta}} \left\{ \int_{f(\mathbf{v})} E(Y) N(f(\mathbf{v}) | \tilde{\theta}, J) df(\mathbf{v}) \right\} N(\tilde{\theta} | m(\mathbf{v}), S(\mathbf{v})) p(\boldsymbol{\psi} | D) d\tilde{\theta} d\boldsymbol{\psi} \\
&= \int_{\boldsymbol{\psi}} \int_{\tilde{\theta}} \left\{ \int_{f(\mathbf{v})} \left\{ \int_{\mathbf{v}} f(\mathbf{v}) dH(\mathbf{v}) \right\} N(f(\mathbf{v}) | \tilde{\theta}, J) df(\mathbf{v}) \right\} N(\tilde{\theta} | m(\mathbf{v}), S(\mathbf{v})) p(\boldsymbol{\psi} | D) d\tilde{\theta} d\boldsymbol{\psi} \\
&= \int_{\boldsymbol{\psi}} \int_{\tilde{\theta}} \left\{ \int_{\mathbf{v}} \left\{ \int_{f(\mathbf{v})} f(\mathbf{v}) N(f(\mathbf{v}) | \tilde{\theta}, J) df(\mathbf{v}) \right\} dH(\mathbf{v}) \right\} N(\tilde{\theta} | m(\mathbf{v}), S(\mathbf{v})) p(\boldsymbol{\psi} | D) d\tilde{\theta} d\boldsymbol{\psi} \\
&= \int_{\boldsymbol{\psi}} \int_{\tilde{\theta}} \left\{ \int_{\mathbf{v}} \tilde{\theta} dH(\mathbf{v}) \right\} N(\tilde{\theta} | m(\mathbf{v}), S(\mathbf{v})) p(\boldsymbol{\psi} | D) d\tilde{\theta} d\boldsymbol{\psi} \\
&= \int_{\boldsymbol{\psi}} \left\{ \int_{\mathbf{v}} m(\mathbf{v}) \prod_{\ell=1}^7 dH_{\ell}(v_{\ell}) \right\} p(\boldsymbol{\psi} | D) d\boldsymbol{\psi} \\
&= \int_{\boldsymbol{\psi}} \left\{ \int_{\mathbf{v}} \left\{ \mu + \mathbf{r}^T(\mathbf{v}) R_{\phi}^{-1} (\boldsymbol{\theta} - \mu \mathbf{1}_n) \right\} \prod_{\ell=1}^7 dH_{\ell}(v_{\ell}) \right\} p(\boldsymbol{\psi} | D) d\boldsymbol{\psi} \\
&= \int_{\boldsymbol{\psi}} \left\{ \mu + \mathbf{T}^T R_{\phi}^{-1} (\boldsymbol{\theta} - \mu \mathbf{1}_n) \right\} p(\boldsymbol{\psi} | D) d\boldsymbol{\psi},
\end{aligned}$$

where \mathbf{T} is the $n \times 1$ vector with i -th element given by

$$\prod_{\ell=1}^7 \left\{ \int_0^1 \exp(-\phi_{\ell} |v_{\ell} - x_{\ell i}|) dv_{\ell} \right\} = \prod_{\ell=1}^7 \left\{ \frac{2 - e^{-\phi_{\ell} x_{\ell i}} - e^{-\phi_{\ell} (1-x_{\ell i})}}{\phi_{\ell}} \right\}$$

For each specified value u_j of the j -th input, we have

$$E(Y | u_j) = \int_{\{v_{\ell}: \ell \neq j\}} f(v_1, \dots, u_j, \dots, v_7) \prod_{\{\ell: \ell \neq j\}} dH_{\ell}(v_{\ell})$$

Let $\mathbf{v}_j \equiv f(v_1, \dots, u_j, \dots, v_7)$, then

$$\begin{aligned}
E^* \{E(Y|u_j) | D\} &= \int_{f(\mathbf{v}_j)} E(Y|u_j) p(f(\mathbf{v}_j)|D) df(\mathbf{v}_j) \\
&= \int_{f(\mathbf{v}_j)} E(Y|u_j) \left\{ \int_{\boldsymbol{\psi}} \int_{\tilde{\theta}} N(f(\mathbf{v}_j)|\tilde{\theta}, J) N(\tilde{\theta}|m(\mathbf{v}_j), S(\mathbf{v}_j)) p(\boldsymbol{\psi}|D) d\tilde{\theta} d\boldsymbol{\psi} \right\} df(\mathbf{v}_j) \\
&= \int_{\boldsymbol{\psi}} \int_{\tilde{\theta}} \left\{ \int_{f(\mathbf{v}_j)} E(Y|u_j) N(f(\mathbf{v}_j)|\tilde{\theta}, J) df(\mathbf{v}_j) \right\} N(\tilde{\theta}|m(\mathbf{v}_j), S(\mathbf{v}_j)) p(\boldsymbol{\psi}|D) d\tilde{\theta} d\boldsymbol{\psi} \\
&= \int_{\boldsymbol{\psi}} \int_{\tilde{\theta}} \left\{ \int_{f(\mathbf{v}_j)} \left\{ \int_{\mathbf{v}} f(\mathbf{v}_j) dH(\mathbf{v}) \right\} N(f(\mathbf{v}_j)|\tilde{\theta}, J) df(\mathbf{v}_j) \right\} \\
&\quad \times N(\tilde{\theta}|m(\mathbf{v}_j), S(\mathbf{v}_j)) p(\boldsymbol{\psi}|D) d\tilde{\theta} d\boldsymbol{\psi} \\
&= \int_{\boldsymbol{\psi}} \int_{\tilde{\theta}} \left\{ \int_{\mathbf{v}} \left\{ \int_{f(\mathbf{v}_j)} f(\mathbf{v}_j) N(f(\mathbf{v}_j)|\tilde{\theta}, J) df(\mathbf{v}_j) \right\} dH(\mathbf{v}) \right\} \\
&\quad \times N(\tilde{\theta}|m(\mathbf{v}_j), S(\mathbf{v}_j)) p(\boldsymbol{\psi}|D) d\tilde{\theta} d\boldsymbol{\psi} \\
&= \int_{\boldsymbol{\psi}} \int_{\tilde{\theta}} \left\{ \int_{\mathbf{v}} \tilde{\theta} dH(\mathbf{v}) \right\} N(\tilde{\theta}|m(\mathbf{v}_j), S(\mathbf{v}_j)) p(\boldsymbol{\psi}|D) d\tilde{\theta} d\boldsymbol{\psi} \\
&= \int_{\boldsymbol{\psi}} \left\{ \int_{\{v_\ell: \ell \neq j\}} m(\mathbf{v}_j) \prod_{\{\ell: \ell \neq j\}} dH_\ell(v_\ell) \right\} p(\boldsymbol{\psi}|D) d\boldsymbol{\psi} \\
&= \int_{\boldsymbol{\psi}} \left\{ \mu + T_j^T(u_j) R_\phi^{-1}(\boldsymbol{\theta} - \mu \mathbf{1}_n) \right\} p(\boldsymbol{\psi}|D) d\boldsymbol{\psi},
\end{aligned}$$

where $T_j(u_j)$ is the $n \times 1$ vector with i -th element given by

$$\exp(-\phi_j |u_j - x_{ji}|) \times \prod_{\ell \neq j} \left\{ \int_0^1 \exp(-\phi_\ell |v_\ell - x_{\ell i}|) dv_\ell \right\}.$$

Denote $\mathbf{v}_j \equiv (v_1, \dots, u_j, \dots, v_7)$ and $\mathbf{v}'_j \equiv (v'_1, \dots, u_j, \dots, v'_7)$. Then,

$$\begin{aligned}
(E(Y|u_j))^2 &= \left(\int_{\{v_\ell: \ell \neq j\}} f(\mathbf{v}_j) \prod_{\{\ell: \ell \neq j\}} dH_\ell(v_\ell) \right)^2 \\
&= \int_{\{v_\ell: \ell \neq j\}} \int_{\{v'_\ell: \ell \neq j\}} f(\mathbf{v}_j) f(\mathbf{v}'_j) \times \prod_{\{\ell: \ell \neq j\}} dH_\ell(v_\ell) \prod_{\{\ell: \ell \neq j\}} dH_\ell(v'_\ell)
\end{aligned}$$

and therefore taking expectation with respect to the bivariate posterior predictive distribution for $(f(\mathbf{v}_j), f(\mathbf{v}'_j))$, developed in Section 3.2, we obtain

$$\begin{aligned}
E^* \left\{ (E(Y|u_j))^2 \mid D \right\} &= \int (E(Y|u_j))^2 p(f(\mathbf{v}_j), f(\mathbf{v}'_j) \mid D) df(\mathbf{v}_j) df(\mathbf{v}'_j) \\
&= \int (E(Y|u_j))^2 \left\{ \int_{\boldsymbol{\psi}} \int_{(\tilde{\theta}, \tilde{\theta}')} N_2 \left(f(\mathbf{v}_j), f(\mathbf{v}'_j) \mid (\tilde{\theta}, \tilde{\theta}'), JI_2 \right) \right. \\
&\quad \left. N_2 \left(\tilde{\theta}, \tilde{\theta}' \mid \boldsymbol{\omega}(\mathbf{v}_j, \mathbf{v}'_j), C(\mathbf{v}_j, \mathbf{v}'_j) \right) p(\boldsymbol{\psi} \mid D) d\tilde{\theta} d\tilde{\theta}' d\boldsymbol{\psi} \right\} df(\mathbf{v}_j) df(\mathbf{v}'_j) \\
&= \int_{\boldsymbol{\psi}} \int_{(\tilde{\theta}, \tilde{\theta}')} \left\{ \int \left\{ \iint_{\substack{\{v_\ell: \ell \neq j\} \\ \{v'_\ell: \ell \neq j\}}} f(\mathbf{v}_j) f(\mathbf{v}'_j) \times \prod_{\{\ell: \ell \neq j\}} dH_\ell(v_\ell) \prod_{\{\ell: \ell \neq j\}} dH_\ell(v'_\ell) \right\} \right. \\
&\quad \left. N_2 \left(f(\mathbf{v}_j), f(\mathbf{v}'_j) \mid (\tilde{\theta}, \tilde{\theta}'), JI_2 \right) df(\mathbf{v}_j) df(\mathbf{v}'_j) \right\} \\
&\quad N_2 \left(\tilde{\theta}, \tilde{\theta}' \mid \boldsymbol{\omega}(\mathbf{v}_j, \mathbf{v}'_j), C(\mathbf{v}_j, \mathbf{v}'_j) \right) p(\boldsymbol{\psi} \mid D) d\tilde{\theta} d\tilde{\theta}' d\boldsymbol{\psi} \\
&= \int_{\boldsymbol{\psi}} \int_{(\tilde{\theta}, \tilde{\theta}')} \left\{ \iint_{\substack{\{v_\ell: \ell \neq j\} \\ \{v'_\ell: \ell \neq j\}}} \left\{ \int f(\mathbf{v}_j) f(\mathbf{v}'_j) N_2 \left(f(\mathbf{v}_j), f(\mathbf{v}'_j) \mid (\tilde{\theta}, \tilde{\theta}'), JI_2 \right) df(\mathbf{v}_j) df(\mathbf{v}'_j) \right\} \right. \\
&\quad \left. \times \prod_{\{\ell: \ell \neq j\}} \prod_{\{\ell: \ell \neq j\}} dH_\ell(v_\ell) dH_\ell(v'_\ell) \right\} N_2 \left(\tilde{\theta}, \tilde{\theta}' \mid \boldsymbol{\omega}(\mathbf{v}_j, \mathbf{v}'_j), C(\mathbf{v}_j, \mathbf{v}'_j) \right) p(\boldsymbol{\psi} \mid D) d\tilde{\theta} d\tilde{\theta}' d\boldsymbol{\psi} \\
&= \int_{\boldsymbol{\psi}} \int_{(\tilde{\theta}, \tilde{\theta}')} \left\{ \iint_{\substack{\{v_\ell: \ell \neq j\} \\ \{v'_\ell: \ell \neq j\}}} \tilde{\theta} \tilde{\theta}' \prod_{\{\ell: \ell \neq j\}} dH_\ell(v_\ell) \prod_{\{\ell: \ell \neq j\}} dH_\ell(v'_\ell) \right\} \\
&\quad N_2 \left(\tilde{\theta}, \tilde{\theta}' \mid \boldsymbol{\omega}(\mathbf{v}_j, \mathbf{v}'_j), C(\mathbf{v}_j, \mathbf{v}'_j) \right) p(\boldsymbol{\psi} \mid D) d\tilde{\theta} d\tilde{\theta}' d\boldsymbol{\psi} \\
&= \int_{\boldsymbol{\psi}} \left\{ \iint_{\substack{\{v_\ell: \ell \neq j\} \\ \{v'_\ell: \ell \neq j\}}} \tilde{E} \left(\tilde{\theta} \tilde{\theta}' \mid \boldsymbol{\psi} \right) \times \prod_{\{\ell: \ell \neq j\}} dH_\ell(v_\ell) \prod_{\{\ell: \ell \neq j\}} dH_\ell(v'_\ell) \right\} p(\boldsymbol{\psi} \mid D) d\boldsymbol{\psi}
\end{aligned} \tag{12}$$

where \tilde{E} denotes expectation with respect to the bivariate normal $N_2 \left(\tilde{\theta}, \tilde{\theta}' \mid \boldsymbol{\omega}(\mathbf{v}_j, \mathbf{v}'_j), C(\mathbf{v}_j, \mathbf{v}'_j) \right)$.

Using the standard covariance identity,

$$\tilde{E} \left(\tilde{\theta} \tilde{\theta}' \mid \boldsymbol{\psi} \right) = \tilde{Cov} \left(\tilde{\theta}, \tilde{\theta}' \mid \boldsymbol{\psi} \right) + \tilde{E} \left(\tilde{\theta} \mid \boldsymbol{\psi} \right) \tilde{E} \left(\tilde{\theta}' \mid \boldsymbol{\psi} \right). \tag{13}$$

Denote $\mathbf{R}_1 \equiv \mathbf{R}_1(v_1, \dots, u_j, \dots, v_7)$ and $\mathbf{R}_2 \equiv \mathbf{R}_2(v'_1, \dots, u_j, \dots, v'_7)$ the first and second columns, respectively, of the $(n \times 2)$ matrix $R(\mathbf{v}, \mathbf{v}')$ defined in Section 3.2. Note that here

the input vectors we are working with, $(v_1, \dots, u_j, \dots, v_7)$ and $(v'_1, \dots, u_j, \dots, v'_7)$, have common element u_j . Therefore, \mathbf{R}_1 is the $n \times 1$ vector with elements

$$\exp \left(-\phi_j |u_j - x_{ij}| - \sum_{\{\ell: \ell \neq j\}} \phi_\ell |v_\ell - x_{i\ell}| \right), \quad i = 1, \dots, n,$$

and analogously for \mathbf{R}_2 , replacing v_ℓ with v'_ℓ . Then using (5) and (6), we obtain

$$\tilde{E}(\tilde{\theta} \mid \boldsymbol{\psi}) = \mu + \mathbf{R}_1^T R_\phi^{-1} (\boldsymbol{\theta} - \mu \mathbf{1}_n) \quad (14)$$

$$\tilde{E}(\tilde{\theta}' \mid \boldsymbol{\psi}) = \mu + \mathbf{R}_2^T R_\phi^{-1} (\boldsymbol{\theta} - \mu \mathbf{1}_n) \quad (15)$$

$$\tilde{Cov}(\tilde{\theta}, \tilde{\theta}' \mid \boldsymbol{\psi}) = \tau^2 \left\{ \exp \left(- \sum_{\{\ell: \ell \neq j\}} \phi_\ell |v_\ell - v'_\ell| \right) - \mathbf{R}_1^T R_\phi^{-1} \mathbf{R}_2 \right\} \quad (16)$$

Substituting (13), (14), (15), and (16), in (12), we obtain for each $j = 1, \dots, 7$,

$$E^* \left\{ (E(Y|u_j))^2 \mid D \right\} = \int_{\boldsymbol{\psi}} \left\{ \tau^2 \left(e - \mathbf{T}_j^T(u_j) R_\phi^{-1} T_j(u_j) \right) + \left(\mu + \mathbf{T}_j^T(u_j) R_\phi^{-1} (\boldsymbol{\theta} - \mu \mathbf{1}_n) \right)^2 \right\} p(\boldsymbol{\psi} \mid D) d\boldsymbol{\psi}$$

where

$$e = \prod_{\{\ell: \ell \neq j\}} \left\{ \int_0^1 \int_0^1 \exp(-\phi_\ell |v_\ell - v'_\ell|) dv_\ell dv'_\ell \right\} = \prod_{\{\ell: \ell \neq j\}} \frac{2}{\phi_j^2} (e^{-\phi_j} + \phi_j - 1)$$

APPENDIX C

Here, we provide the details for generating the distributions of the main effects and sensitivity indices.

We begin by generating input sample matrix \mathbf{M} of size $B \times 7$,

$$\mathbf{M} = \begin{pmatrix} v_{1,1} & v_{1,2} & \dots & v_{1,7} \\ v_{2,1} & v_{2,2} & \dots & v_{2,7} \\ \vdots & \vdots & \dots & \vdots \\ v_{B,1} & v_{B,2} & \dots & v_{B,7} \end{pmatrix}$$

where each of its elements is a random sample from $\text{Unif}(0, 1)$, or equivalently, each row of \mathbf{M} is a random sample from $P(v_1, \dots, v_7)$.

Next, we generate 7 input sample matrices, \mathbf{N}_j , for $j = 1, \dots, 7$, of size $B \times 7$ each,

$$\mathbf{N}_j = \begin{pmatrix} v'_{1,1} & v'_{1,2} & \dots & v_{1,j} & \dots & v'_{1,7} \\ v'_{2,1} & v'_{2,2} & \dots & v_{2,j} & \dots & v'_{2,7} \\ \vdots & \vdots & \dots & \vdots & \dots & \vdots \\ v'_{B,1} & v'_{B,2} & \dots & v_{B,j} & \dots & v'_{B,7} \end{pmatrix}$$

where the j -th column of matrix \mathbf{N}_j equals the j -th column of matrix \mathbf{M} , but the elements of the other 6 columns are randomly sampled from $\text{Unif}(0, 1)$.

Then, we generate 7 input sample matrices, \mathbf{N}_{-j} , for $j = 1, \dots, 7$, of size $B \times 7$ each,

$$\mathbf{N}_{-j} = \begin{pmatrix} v_{1,1} & v_{1,2} & \dots & v'_{1,j} & \dots & v_{1,7} \\ v_{2,1} & v_{2,2} & \dots & v'_{2,j} & \dots & v_{2,7} \\ \vdots & \vdots & \dots & \vdots & \dots & \vdots \\ v_{B,1} & v_{B,2} & \dots & v'_{B,j} & \dots & v_{B,7} \end{pmatrix}$$

where the $(-j)$ -th columns of matrix \mathbf{N}_{-j} equal the $(-j)$ -th columns of matrix \mathbf{M} , but the elements of the j column are randomly sampled from $\text{Unif}(0, 1)$.

Then, we generate vector $\mathbf{w} = \{w_k : k = 1, 2, \dots, 50\}$, where the elements w_k are equally spaced between 0 and 1.

And finally, we generate 350 matrices $\mathbf{M}_j^{(k)}$ such that for every $j = 1, \dots, 7$, and for every $k = 1, \dots, 50$, each element of the j -th column in \mathbf{M} is equal to w_k .

$$\mathbf{M}_j^{(k)} = \begin{pmatrix} v_{1,1} & v_{1,2} & \dots & v_{1,j} = w_k & \dots & v_{1,7} \\ v_{2,1} & v_{2,2} & \dots & v_{2,j} = w_k & \dots & v_{2,7} \\ \vdots & \vdots & \dots & \vdots & \dots & \vdots \\ v_{B,1} & v_{B,2} & \dots & v_{B,j} = w_k & \dots & v_{B,7} \end{pmatrix}$$

Now, we are ready to start the Monte Carlo simulation.

1. Start at current MCMC state, $\boldsymbol{\psi}^{(t)} = \left(\boldsymbol{\theta}^{(t)}, \mu^{(t)}, \tau^{2(t)}, \phi_1^{(t)}, \phi_2^{(t)}, \dots, \phi_7^{(t)} \right)$, and compute the observed correlation matrix, $R_{\boldsymbol{\phi}^{(t)}}^{-1}$.
 - (a) For each row i of M , sample $\tilde{y}_i \equiv \tilde{y}_i^{(t)}$ according to (2), then compute \tilde{y}_i^2 .
 - (b) For each row i of N_j , sample $\tilde{y}'_{i,j}$ according to (2).
 - (c) For each row i of N_{-j} , sample $\tilde{y}'_{i,-j}$ according to (2).
 - (d) For each row i of $M_j^{(k)}$, sample $\tilde{y}_{i,j}^{(k)}$ according to (2).
2. Obtain the t -th posterior sample from $P(E(Y)|D)$ and the t -th posterior sample from $P(E(Y^2)|D)$,

$$E(Y)^{(t)} = \frac{1}{B} \sum_{i=1}^B \tilde{y}_i$$

$$E(Y^2)^{(t)} = \frac{1}{B} \sum_{i=1}^B \tilde{y}_i^2$$

3. For $j = 1, \dots, 7$, obtain the t -th posterior sample from $P\left(E\left((E(Y|u_j))^2\right) | D\right)$,

$$E\left((E(Y|u_j))^2\right)^{(t)} = \frac{1}{B-1} \sum_{i=1}^B \tilde{y}_i \tilde{y}'_{i,j}$$

4. For $j = 1, \dots, 7$, obtain the t -th posterior sample from $P\left(E\left((E(Y|u_{-j}))^2\right)|D\right)$,

$$E\left((E(Y|u_{-j}))^2\right)^{(t)} = \frac{1}{B-1} \sum_{i=1}^B \tilde{y}_i \tilde{y}'_{i,-j}$$

5. For each $k = 1, \dots, 50$, and for each $j = 1, \dots, 7$, obtain the t -th posterior sample from $P(E(Y|v_j = w_k)|D)$,

$$E(Y|v_j = w_k)^{(t)} = \frac{1}{B} \sum_{i=1}^B \tilde{y}_{i,j}^{(k)}$$

6. Obtain $S_j^{(t)}$ according to (9).

7. Obtain $S_j^T(t)$ according to (10).

8. For each $k = 1, \dots, 50$, and for each $j = 1, \dots, 7$, obtain $f_j^{(t)}$ according to (8).

Retrieval of Vegetation Canopy Biophysical Parameters via Remote Sensing and Radiative Transfer Models: An Overview

Roberto Furfaro^a and Barry D. Ganapol^{a,b}

^aAerospace and Mechanical Engineering Department

^bHydrology and Water Resources

University of Arizona, Tucson, Arizona

Corresponding Author: Roberto Furfaro, Assistant Research Professor,
email:robertof@email.arizona.edu, tel: 001-520-312-7440

Abstract

Estimating canopy biophysical parameters from data collected by sensors mounted on-board Earth orbiting spacecrafts is one of the fundamental problems in remote sensing. Operational retrieval algorithms have been developed to provide the most accurate estimation of Leaf Area Index (LAI) and fraction of Absorbed Photosynthetically Active Radiation (fAPAR) on a global scale, since those parameters control the exchange of carbon between atmosphere and vegetated surface. In this paper, an overview of the methodology currently used in designing inverse algorithms for canopy biophysical parameters retrieval is presented. The most efficient algorithms have been devised by inverting canopy Radiative Transfer (RT) models. Understanding the physics of the interaction between photons and vegetation is a critical component of the overall procedure. Canopy RT theory is discussed first to provide an understanding of the origin of the physically-based models that have been employed during the estimation process. After discussing some of the major issues arising in model-based inversion (e.g. ill-posed nature of the inversion, model sensitivity, optimal sampling), three major inverse techniques, i.e. traditional iterative optimization, look-up tables and neural networks, are presented. The current state-of-the-art for the practice of developing retrieval algorithms, with special emphasis on LAI operational algorithms, is discussed. While traditional iterative techniques are useful to illustrate the basics of the inversion, a review of the literature shows that neural network algorithms are gaining increasing popularity and seem to provide an ideal solution for the canopy biophysical parameters retrieval problem. However, the look-up table inversion remains the most intuitive and simple algorithm and it is currently the basis for the major operational LAI retrieval algorithm currently available. Validation and intercomparison between various satellite-based LAI products and algorithms is reviewed. Finally, the operational characteristic of four major satellite sensors (MODIS, MISR, VEGETATION and MERIS) commonly used for vegetation monitoring, are discussed to provide a link between measurements and retrieval algorithms.

1. Introduction

Vegetation canopy biophysical parameters play an important role in controlling land surface processes. Large-scale ecosystem models developed to simulate how the Earth system responds to changing in climate and atmospheric composition, require accurate knowledge and quantification of canopy structural and biochemical parameters. For example, all process models

capable of simulating the exchange of carbon between atmosphere and vegetated surfaces need Leaf Area Index (LAI) and fraction of Absorbed Photosynthetically Active Radiation (fAPAR) as state variables (Sellers et. al. 1992). Clearly, vegetation canopies have a critical impact on both global carbon cycle and surface-atmosphere energy exchange. Most importantly, processes such as canopy evapotranspiration and photosynthesis can be linked to canopy spectral reflectance. The electromagnetic energy reflected by vegetation canopies in any direction and at any specified wavelength results from a plethora of complex physiological, biochemical and physical processes. It is characterized by morphological spatial and temporal canopy variations. Satellite remote sensing provides a unique way to aid in quantifying the required vegetation parameters and in monitoring the state of vegetation canopies around the world. Many satellite-based remote sensors have been deployed in Earth's orbit to provide a comprehensive sampling of the electromagnetic spectrum reflected by the vegetative surfaces. Since most ecosystem models do not use the reflected radiation directly, the collected data must be processed by retrieval algorithms designed and implemented to estimate the desired canopy biophysical parameters. The design and implementation of retrieval algorithms can be defined as the intermediate step necessary to transform the collected radiation into estimated canopy structural and biochemical parameters. Generally, two broad classes of methods are available to implement the data, i.e. empirical methods and physically-based models.

Empirical methods deliver the desired canopy biophysical variables using semi-empirical relationships connecting the variable of interest (e.g. LAI, chlorophyll) to radiometric information (Vegetation Indices (VI), see Buermann et. al. 2002). The VI-based approach is computationally efficient especially when processing large amounts of data (e.g. the relationship between VIs and LAI is generally expressed in analytical form) and can give satisfactory results under favorable conditions (Atzberger et. al. 2003). However, they lack portability, i.e. they are highly specialized for specific sensors and sites (Gobron et al. 1997). Indeed, since the canopy spectral and directional reflectance depends on many parameters and conditions, it is extremely hard to find a unique, one-to-one relationship between radiometric information and desired canopy biophysical variables that is applicable at any spatial and temporal scale (Currant 1994).

Physically-based, first-principle, models may serve as a basis for extracting vegetation parameters. They rely on heavy usage of Radiative Transfer (RT) models (Ganapol et. al. 1999, 2006) and over the past two decades they lead the way to the development of more reliable and efficient algorithms for canopy biophysical parameters estimation. RT models describe the physical processes occurring when photons interact with canopy phytoelements such as leaves, stems, barks, as well as soil. They account for the transport of photons within vegetation structures both directionally and spectrally. Canopy RT models are able to simulate the electromagnetic energy reflected by canopies as a function of their structural (e.g. amount of one-sided green leaf area or LAI) and biochemical parameters (e.g. chlorophyll and water content). Thus, RT models are flexible tools that enable scientists to exploit the full radiometric information collected by remote sensors. However, devising an operational RT model-based retrieval algorithm is not a simple process as it requires a model inversion, i.e. given the collected radiometric information, the canopy RT model is used to match simulated observations with measured observations (Verstraete et. al. 1996).

This review deals with both canopy radiative transfer models and their inversion. Since retrieval algorithms are intimately connected with the inverted RT models, an overview of both the radiative transfer modeling process and the inversion techniques are appropriate. We will review the fundamentals of both sides of the process, spanning from the physics of the interaction between photons and vegetation to the numerical inversion techniques. A review of the state-of-the-art in operational algorithms for canopy biophysical parameters estimation with a special emphasis on LAI is presented, including implementation, performance analysis and validation.

Our paper is organized as follows. Section 2 presents a brief statement of the fundamental problem in remote sensing-- modeling and inversion. Section 3 reviews the basics behind radiative transfer in vegetation canopies. Few examples of available one-dimensional and three-dimensional canopy RT models commonly used for inversion are then presented. Section 4 reviews the fundamentals of RT model inversion and presents three basic techniques, i.e. iterative optimization, look-up- tables and neural networks. A review of the state-of-the-art of inversion algorithms, their current implementation, validation and performance is presented with special emphasis on operational algorithms for LAI retrieval. Section 5 briefly reviews the characteristics of satellite-based sensors commonly employed in observing vegetation canopies. Section 6 draws the conclusions.

2. A Fundamental Problem in Remote Sensing: Canopy Biophysical Parameters Retrieval

From remote sensing we infer the physical properties of a body by measuring electromagnetic radiation either emitted or reflected from an independent source (Schott 1997). In the case of vegetation canopies, their biophysical parameters are causally connected to the radiance measured by observing sensors. Physically-based canopy radiative transfer models (sometimes also called canopy reflectance (CR) models) can be used to model the direct relationship between radiative state variable and observed measurements. A formal mathematical representation of the direct (or “forward”) relationship can be obtained as follows (Combal et. al. 2002):

$$\mathbf{R} = CRTmodel(\mathbf{P}_V, \mathbf{P}_C) + \varepsilon \quad (1)$$

Equation (1) is the functional representation of the direct/forward canopy problem where the response to passive illumination of a canopy realization defined by N vegetation parameters $\mathbf{P}_V = \{ P_{V1}, P_{V2}, \dots, P_{VN} \}$ and M configuration parameters $\mathbf{P}_C = \{ P_{C1}, P_{C2}, \dots, P_{CM} \}$ (e.g., sun illumination, sensor angles, wavelengths), is simulated via a canopy RT model to match the observed reflectances \mathbf{R} . Here, ε represents the discrepancy between observed and modeled data. Two sources are responsible for the mismatch, i.e. measurement error and model uncertainty. Measurement errors, an unavoidable component of the measurement process, arise due to sensor noise and data processing which is required to transform raw sensor data in physical quantities simulated by the model. Modeling error/uncertainty, also unavoidable, is generated by the assumptions made when modeling the physical interaction between photons and vegetation.

Retrieving canopy biophysical parameters from measurements and RT models, can be approached by looking for a “match” between modeled and observed reflectances. This requires the solution of an inverse problem, where we are given the measurements and we are to

determine the parameters of the observed surface. Figure (1) provides a schematic of process. Solving the inversion problem is intimately connected with the nature of the RT model employed to describe the radiative field within and at the top of the canopy and requires a thorough understanding of the canopy RT model behavior as a function of the canopy parameters.

In the following two sections, we review the fundamentals and state-of-the-art in both canopy RT modeling and the inversion techniques required to derive a RT-based retrieval algorithm.

3. Physically-based first principle models for Radiative Transfer in Vegetation Canopies

Physically-based, first-principle, models mathematically describe the physical processes behind the interaction between electromagnetic radiation and vegetation. They are critical to our understanding, both quantitatively and qualitatively, of the nature of the signal collected by satellite-based sensors. While a large variety of approaches have been taken (e.g. geometric-optical (Li and Strahler, 1992), radiosity (Lewis, 1999), Ray-tracing and Monte Carlo (Govaert and Verstraete, 1998, North, 1996), here we focus on physical models that employ radiative transfer theory to describe the transport of photons in vegetation. The rationale behind considering primarily RT models is that most operational algorithms currently employed in estimating vegetation biophysical parameters (both at local and global scale), are based on their inversion (e.g. Baret et al. 2007). Since RT models describe the transport of photons in a host medium, they are generally derived by applying the principle of photons. Canopies exhibit extremely rich and complex architectures and therefore several assumptions are generally required to establish mathematically tractable models. In the RT theory of vegetation canopies, the canopy is idealized as a medium filled with infinitesimally small planar elements, i.e. plates of negligible thickness oriented according to a specified probability law. All phytoelements other than green leaves (e.g. stems, barks) are ignored. Overall, vegetation canopies are viewed as a gas (or cloud) of non-dimensional planar scattering centers where the finite size of the leaves (the primary scatterers) is neglected. Therefore, the “turbid medium” assumption (Ross, 1981) describes the canopy as a collection of independent dimensionless scatterers uniformly distributed within the medium whose distribution and orientation called canopy architecture are described by leaf area density and leaf normal distribution functions. Those functions convey information about the total amount of green leaf and the leaf orientation within the canopy. In the remainder of the section, we provide a relatively brief overview of the classical canopy RT theory as developed by Ross (1981) and described by Shultis and Myneni (1989), Ross and Myneni (1991) and Knyazkhin et al. (2005). Subsequently, three examples of popular canopy RT models employed in deriving model-based inversion algorithms are reviewed.

3.1 Fundamentals of canopy RT theory: Optical models and RT equations.

Estimation of the radiative regime within vegetation canopies bounded by a reflecting soil requires the definition of 1) the architecture of individual plants and full canopy; and 2) definition of the optical properties of leaves (scatterers) and soil (as a boundary condition).

Two wavelength-independent structural parameters define the architecture of canopy, i.e. the leaf area density and leaf normal orientation distribution functions. The leaf area density distribution function $u_L(\mathbf{r})$ is defined as the one-sided green leaf area per unit volume and it is a function of

the position inside the canopy $\mathbf{r} = (x, y, z)$. Integrating $u_L(\mathbf{r})$ along the entire canopy height H , we obtain the Leaf Area Index (LAI):

$$LAI(x, y) = \int_0^H u_L(x, y, z) dz \quad (2)$$

LAI is defined as the one-sided green leaf area per unit ground area and is a function of the horizontal position (x, y) within the observed canopy.

If $\mathbf{\Omega}_L(\varphi_L, \psi_L)$ is the upward normal to the leaf element (φ_L is the leaf inclination angle and ψ_L is the leaf azimuthal angle), we define $(1/2\pi)g_L(\mathbf{\Omega}_L)$ to be the probability density distribution of the leaf normal distribution with respect to the upper hemisphere (also known as Leaf Angle Distribution or LAD). The function g_L is commonly described as a product of two independent functions, one of the inclination angle and the other of azimuthal angle only (Shultis and Myneni 1989). Ample evidence indicates that the leaves are randomly oriented in the azimuthal angle whereas they follow five possible general distributions in the inclination angle, i.e. planophile, plagiofile, erectophile, extremophile and uniform (DeWit 1965, Bunnick 1978). For example, planophile indicates a class of leaves mainly distribute horizontally, while erectophile defines a class of leaves mainly distributed vertically.

From an optical point of view, single-leaf reflectance and transmittance are required to define the so-called leaf scattering phase function (Ross 1981, Shultis and Myneni 1989, Myneni et. al. 1990, Pinty and Verstrate 1997, Knyazkhin et al., 2005). Symbolically represented as $\varphi_L(\mathbf{\Omega}', \mathbf{\Omega}; \mathbf{\Omega}_L)$, it models the probability that a photon travelling in direction $\mathbf{\Omega}'$ is scattered by a leaf with normal $\mathbf{\Omega}_L$ in direction $\mathbf{\Omega}$. The mathematical form of the leaf scattering phase function depends on the model used to describe the interaction between photons and leaves. The most popular model assumes the leaf to behave as a bi-lambertian diffuser (Shultis and Myneni 1989). Specular reflectance models are also available (Vanderbilt et. al. 1991).

The conventional transport of photons in a participating medium requires the definition of extinction, absorption and scattering coefficients (Chandrasekhar 1950). Such coefficients are intimately connected with the physical processes occurring within the medium. In the case of vegetation canopy, photon absorption and scattering depend on the canopy structural parameters. The extinction coefficient is defined as the probability, per unit path length, of a photon travelling in direction $\mathbf{\Omega}$ to encounter a leaf element. Mathematically,

$$\sigma(\mathbf{r}, \mathbf{\Omega}) = u_L(\mathbf{r})G(\mathbf{r}, \mathbf{\Omega}) = \frac{1}{2\pi} \int_+ d\mathbf{\Omega}_L g_L(\mathbf{r}, \mathbf{\Omega}_L) |\mathbf{\Omega} \cdot \mathbf{\Omega}_L| \quad (3)$$

The $G(\mathbf{r}, \mathbf{\Omega})$ function is called geometric function (Ross, 1981, Shultis and Myneni 1989) and it is defined as the projection of a unit leaf area located at \mathbf{r} onto a plane perpendicular to the direction of photons travel $\mathbf{\Omega}$.

The differential scattering coefficient is defined as the probability density, per unit path length, that a photon, impinging on a leaf element of area density $u_L(\mathbf{r})$ in direction $\mathbf{\Omega}'$, is intercepted and therefore scattered in direction $\mathbf{\Omega}$. Mathematically:

$$\sigma'_{s,\lambda}(\mathbf{r}, \Omega' \rightarrow \Omega) = u_L(\mathbf{r}) \frac{1}{\pi} \Gamma_\lambda(\mathbf{r}, \Omega' \rightarrow \Omega) = u_L(\mathbf{r}) \frac{1}{2\pi} \int_{+} d\Omega_L g_L(\mathbf{r}, \Omega_L) |\Omega \cdot \Omega_L| \gamma_{L,\lambda}(\Omega' \rightarrow \Omega, \Omega_L)$$

(3)

Here, $(1/\pi) \Gamma_\lambda(\mathbf{r}, \Omega' \rightarrow \Omega)$ is the area scattering phase function (Ross 1981, Shultis and Myneni 1989). Scattering and absorption coefficients as well as single scattering albedo can be derived via manipulation of the scattering phase function (for further details see Knyazikhin et al. 2005). The radiative transfer equation can be formulated as a balance of photons in the appropriate phase space. A comprehensive and accurate description of the radiative regime within vegetation canopies requires the solution of a three-dimensional radiative transfer equation. Assuming that the canopy is contained within a finite volume V_C (bounded on the top by the atmosphere ∂V_t , on the bottom by the reflecting soil ∂V_s and laterally by a generic surface ∂V_l), the steady-state radiative transfer equation is formulated as follows (Ross 1981, Shultis and Myneni 1989, Myneni et al. 1991, Knyazikhin et al. 1996, Knyazikhin et al. 2005):

$$\Omega \cdot \nabla I_\lambda(\mathbf{r}, \Omega) + G(\mathbf{r}, \Omega) u_L(\mathbf{r}) I_\lambda(\mathbf{r}, \Omega) = \frac{u_L(\mathbf{r})}{\pi} \int_{4\pi} \Gamma_\lambda(\mathbf{r}, \Omega' \rightarrow \Omega) I_\lambda(\mathbf{r}, \Omega') d\Omega'$$

(4)

Here $I_\lambda(\mathbf{r}, \Omega)$ is the monochromatic (single wavelength) radiation intensity within the vegetation canopy. The canopy equation has been derived assuming no canopy emission and no polarization. Equation (4) is an integro-differential equation (also called the linear Boltzmann equation, Ganapol et. al. 1999). The equation must be augmented with appropriate boundary conditions (BCs). For any RT equation, the BCs represent a statement of the radiant energy entering the medium (Chandrasekhar 1950, Case and Zweifel 1967). In the case of vegetation canopies, photons enter the top of the canopy either via direct illumination of sun or diffuse illumination of the sky (i.e. photons scattered by the atmosphere and redirected into the vegetation). Conversely, at the bottom of the canopy, photons re-enter the medium after being partially reflected by the soil. Any of these conditions can be expressed mathematically (for details see Shultis and Myneni 1989, Kranigk et al. 1994, Knyazikhin et al. 1997, Knyazikhin et al. 2005).

The radiation intensity depends, for any given wavelength, on three spatial and two angular parameters. Because the leaf scatterers are found not to produce any frequency shifting, the radiative field inside canopies can be found independently at any wavelength. The full three-dimensional model can be simplified by assuming radiation fields to depend only on the vertical spatial variable (canopy infinite in the transverse plane) with two angles or one angle (azimuthally integrated intensity) variation. Moreover, in many applications, the canopy is assumed to be homogeneous. There is always a trade-off between the ability of the model to represent rich and complex heterogeneous canopy realizations and reduced mathematical complexity (Verstraete et. al. 1996)

Vegetation canopy RT models cannot be solved analytically and require numerical methods even in the simplified one-dimensional case. Numerical techniques borrowed from the neutron transport field (Bell and Glasstone, 1970) are used to solve vegetation canopy equations. Discrete ordinate methods, where both the spatial and angular variables are discretized, have been applied for the 1-D (Myneni et. al. 1991a) and 3-D case (Myneni 1991b). Accurate and fast numerical solutions have been found applying F_n method (Ganapol and Myneni 1992, Ganapol et al. 1999) which approximate the solution using a singular integral formulation of the transport equation. However, only recently, this highly efficient and accurate algorithm was put on a solid theoretical framework (Furfaro and Ganapol 2007a). More accurate solutions of the canopy RT equation have been obtained by applying the Converged SN method (Ganapol 2006) where both spatial and angular variables are discretized and coupled with acceleration techniques (e.g. Wynn-Epsilon, Sidi 2003) to push the order of accuracy above the sixth decimal place.

Since the RT equation (4) is derived applying rigorous first principles, it satisfies energy conservation. However, the turbid medium approximation fails to model the retro-illumination effect observed commonly by satellite sensors. This effect is called “Hot-Spot” (Kuusk 1991b) and it is observed as a sharp peak of the radiance whenever a vegetation field is observed in the retro-solar direction. This is due to the finite size of the leaves and their ability to cast a shadow. This effect can be approximately modeled by augmenting the first scattered component of the radiation with parameters that empirically describes the peak. Under this conditions, the turbid medium assumption is violated and so is energy conservation. Recently, Kellel et. al. (2008) proposed a new “multi hot-spot” model applied to a four-flux approximation of the canopy RT equation that is able to account for the effect of the finite size of leaves on multiple scattering maintaining energy conservation.

3.2 Advancements in Vegetation Canopy Transport Theory

Three-dimensional canopy RT models are the most comprehensive models to faithfully describe the full radiative field in vegetation. However, they are also computationally expensive. Conversely, one-dimensional models are simpler and require less computational resources, but fail to account for heterogeneity in canopies, and are therefore unable to model leaf clumping and other effects typical of the natural environment. Recently, stochastic transport in vegetation media has been considered as a possible representation of three-dimensional effect using a simpler one-dimensional model (Huang et al. 2008). Stochastic transport theory is not a new concept (Pomraming 1991, 1996) but it has been recently applied to the problem of describing photons transport in discontinuous canopies (Shabanov et al. 2000). The idea is to provide a complete description of the radiative regime in 3-D canopies using a one-dimensional equation resulting from determining the mean radiant field coming from 3-D canopy realizations. This requires the determination of pair-correlation functions using Monte Carlo methods (Huang et al. 2008). However, ignoring the full effect of 3-D canopy structures may result in underestimation of canopy transmittance and consequently overestimation of reflectance and absorptance. Stochastic RT equations seem to adequately account for the effect of canopy structure on relationships between LAI, LAD and reflected energy. While some of the results have been partially implemented for LAI and fAPAR estimations from MODIS data (Shabanov et al. 2005), in the future, it is expected that canopy stochastic theory will provide a better representation of the radiative field resulting in more efficient and reliable retrieval algorithms.

3.3. Popular RT models employed in Canopy Biophysical Estimation Problem

A wide range of radiative transfer models exist in the literature. Here, we review some of the most popular and widely employed in designing operational algorithms for canopy biophysical parameters estimation.

One of the most widely used turbid medium RT models is the Scattering from Arbitrary Inclined Leaves (SAIL) model (Verhoef 1984, 1985). The model, which is an evolution of the Suits model (1972), is able to compute the Bidirectional Reflectance Distribution Function (BRDF) by approximating the monochromatic radiative intensity using a two-flux diffusive approximation to capture multiple scattering between vegetation elements. The multiply scattered fluxes are assumed to be semi-isotropic. Over the past two decades, many modifications and improvements have been proposed including the Hot-Spot implementation (Kuusk 1985, 1991b), yielding the SAILH version (Verhoef 1998), and a two-layer canopy structure to model the gradient color of leaves, yielding GeoSAIL (Verhoef and Bach 2003). In its latest version 4SAIL2 (Verhoef et al. 2007), the model has been upgraded to implement a four-stream diffusive approximation. Moreover, it has the ability to handle soils with anisotropic reflectance. SAIL is often coupled with the radiative transfer model PROSPECT (Jacquemoud and Baret 1990) specifically designed to describe the leaf optical properties as a function of leaf structure and biochemistry. PROSPECT is able to simulate leaf reflectance and transmittance from visible (0.4 microns) to the middle infrared (2.5 microns) as function of a parameter that account for the internal structure, chlorophyll concentration and leaf water depth. The combined model is called SAIL+PROSPECT (or sometimes PROSAIL): it has been used in many model inversion studies and it is the basis for constructing the operational algorithm to retrieve LAI, fAPAR and fraction of vegetation cover from the European VEGETATION sensor (Baret et al. 2007, see also section 4.6).

The coupled Leaf-Canopy RT model (Ganapol et al. 1999) is the only model in the literature that describes the transport of photons both at leaf and canopy levels. Thus, LCM is able to compute the spectral and directional reflectance factor as function of leaf biochemistry, canopy architecture and soil reflectance. The modeling is executed by connecting two radiative transfer models. At leaf level, LEAFMOD (Ganapol et al. 1998) describes the transport of photons within a leafy medium which is assumed to be homogeneous and isotropic. A conventional RT transfer equation is solved to determine the leaf optical properties (i.e. reflectance and transmittance). At canopy level, CANMOD (Ganapol et al. 1999, Ganapol and Myneni 1992) treats the vegetation as a one-dimensional turbid medium. It accepts inputs of the single-leaf optical properties from LEAFMOD as well as LAD, LAI, soil reflectance, sun and view angles to compute the spectral and directional reflectance at the top of the canopy. The leaf model can be calibrated using a combination of leaf experimental optical data and model inversion: usually, for a given leaf-type, LEAFMOD is inverted to compute scattering and absorption coefficients from the experimentally known reflectance and transmittance of a reference leaf. While scattering is kept unchanged as representative of the leaf structure, the retrieved absorption coefficient is disregarded and replaced by a new absorption coefficient constructed from the biochemistry of any desired leaf. Examples of canopy spectral and directional reflectance modeled using LCM is found in figure (2).

LCM has been successfully used in retrieving plantation coffee biophysical parameters from airborne remote sensing data for precision agriculture (Furfaro et al. 2005, 2007b) and it is also employed in a novel approach for global sensitivity analysis of canopy RT models and uncertainty analysis (Morris et. al. 2008).

The Discrete Anisotropic Radiative Transfer (DART) model (Gastellu-Etcheberry et al. 1996) is one of the most complex three-dimensional RT model available in the literature. It was designed to simulate the radiative transfer in three-dimensional, heterogeneous landscape scenes containing multiple elements such as grass, trees, shrubs, soil etc. The general approach to determine the full-scale radiant energy distribution of the scene is to divide the domain in rectangular cells containing single or a combination of elements. Discrete ordinates and kernel techniques have been employed to model the spectral and directional response of a specified observing sensor. DART has been recently improved (Gastellu-Etcheberry et al. 2004) to extend its domain of application and improve accuracy especially regarding simulations of single and multiple scattering. A current DART version is able to simulate the radiative transfer of the whole atmosphere-Earth system including models for sensor transfer functions. Such complex models require a large number of parameters and are computationally expensive. DART is generally used in forward modeling. A limited number of studies have been considered for inversion (e.g. Kimes et al. 2002).

Other canopy RT models with different levels of complexity can be found in the literature (e.g. Markov Chain Reflectance Model, MCRM (Kuusk 1995, 2001), RayTran (Montecarlo simulation, Goevaert and Vestraete 1998)). Recently, the RAdiation transfer Model Intercomparison (RAMI) database (Pinty et. al. 2004, Widlowski et. al. 2007) proposed several protocols to compare vegetation canopy RT models. Such an effort should lead significant advancements in modeling and data interpretation leading to benefits for both the RT modelers and the user communities.

4. Retrieval of biophysical parameters in vegetation canopies: Model Inversion techniques

Inverting canopy RT models requires specific mathematical procedures to find the biophysical parameters that provide the best match between modeled and observed data. In order to invert a canopy RT model against a set of M remote sensing measurements to retrieve P vegetation parameters, Verstrate et.al. (1996) defined four basic conditions that must be satisfied, i.e. 1) the model must be mathematically “well behaved” and must have, as inputs, as less parameters as possible, 2) a cost function must be defined, 3) an inversion procedure that finds the absolute minimum (extremal) of the cost function must exist and 4) more measurements than model parameters must be collected by the sensors.

Clearly, an inversion procedure must be defined to construct an operational algorithm for vegetation biophysical parameters retrieval. The traditional approach to solve the problem is to define an appropriate cost function (also called merit function or objective function) and search the parameter space for the solution(s) that minimize it. Importantly, the goal is not to look for a “true” solution but for an “approximate” solution to the system of equations described in equation (1) in the “best” possible fashion, i.e. solutions that minimize the selected cost function. Formally, equation (1) represents a system of non-linear equations. A necessary condition for the

1 system to be solvable is overdeterminacy ($M > P$, Verstrate et. al. 1996, Goel 1987). Moreover,
2 for a successful implementation of inversion procedures, the RT model linking observations and
3 canopy parameters must be described by the least amount of free parameters. Ideally, one wants
4 to deal with models that have high “variance”, i.e. can represent or “explain” many observations
5 with a very limited number of parameters. Indeed, if a dataset is extremely rich, we can always
6 explain its richness (i.e. large data variance) by allowing models with large number of
7 parameters. However, for these kind of models, the inversion algorithms must search higher-
8 dimensional parameter spaces. Thus, they tend to be unstable, computationally expensing and
9 less reliable.

10 Different inversion strategies have been proposed. We group the inversion algorithms in three
11 categories: a) iterative optimization methods (e.g. Bicheron and Leroy, 1999; Goel and
12 Thompson, 1984; Jacquemoud et al., 2000, Meroni et. al., 2004), b) Look-Up-Tables (LUTs)
13 (e.g. Combal et al., 2002; Knyazikhin et al., 1998a,b; Weiss et al., 2000) and c) Neural Network
14 (NN) inversion algorithms (Baret et. al., 1995; Weiss & Baret, 1999, Bacour et. al. 2006, Baret
15 et. al. 2007). Each approach comes with advantages and disadvantages which are the topic of the
16 remainder of the section. However, before delving into algorithm details, performances and
17 validation, the ill-posed nature of the inversion procedure must be discussed.

18 19 **4.1 Inversion as Ill-posed problem: Regularization**

20
21 From a mathematical point of view, a problem can be solved if and only if it is well-posed (in the
22 sense of Hardamand), i.e. a solution to the problem exists and it depends continuously from the
23 data (Habermann, 2003). However, canopy inverse problems are notoriously ill-posed because a)
24 different model parametrizations may yield identical spectra and b) the presence of model and
25 measurement errors and uncertainties (Combal et. al. 2002, Atzeberger 2004). Baret and Guyot
26 (1991) showed that two canopy realizations with different attributes (dense erectophile and
27 sparse planophile) may yield identical spectral reflectance. Thus, one dataset of observations
28 may yield two different solutions. Moreover, the presence of modeling and measurement errors
29 may induce large variation in the solution, i.e. the “approximate” solution found via the
30 inversion algorithm may not be centered around the “true” solution, but dispersed in the
31 parameter space (Atzberger 2004). Jaquemoud and Baret (1993) used the SAILH+PROSPECT
32 canopy reflectance model to verify the uniqueness of the solution, i.e. they studied inversion
33 feasibility. What was found implied that the solution is not unique (e.g. multiple and different
34 canopy configurations can be found that exhibit similar spectra). Therefore, the successful
35 implementation of inversion algorithms requires the application of some regularization technique
36 to obtain a stable and reliable solution.

37 Regularization can be defined as a technique that aims at selecting only approximate solutions
38 that are closed to the true solution. Combal et. al. (2002) recognized that a-priori information is
39 an effective way to regularize ill-posed problems and conducted a study on using a-priori
40 information to reduce the uncertainty in estimating canopy biophysical parameters. Using a-
41 priori information means acquiring knowledge about the observed field using a) ancillary and
42 field data (i.e. coming from other sensors), b) knowledge of the type of canopy architecture that
43 defines the very specific class of RT models to be employed (e.g. turbid medium) and c)
44 knowledge of typical distributions of the desired biophysical parameters. Combal et. al. (2000)
45 adapted LUT, Neural Networks and Quasi-Newton inversion techniques to account for a-priori
46 information. The study allowed for simulated canopy reflectance data that were adapted to

include model uncertainty and measurement data. They found that while a-priori information always allows the stabilization of the inversion algorithm, LUT and Quasi-Newton algorithms were more sensitive to model uncertainty. However, Neural Networks tend to perform worst when measurement errors are introduced.

Atzberger (2004) contributed to the regularization problem by developing a novel approach that takes into account the neighboring radiometric information of the pixel of interest. The approach, called object signature used a set (“window”) of neighboring pixels to calculate 42 descriptive statistical properties that characterize the spectral covariace of the image object. The pixel signature and the image object signature were used on a Neural-Network-based inversion algorithm (trained using the SAILH+PROSPECT model) capable of retrieving simultaneously three parameters, i.e. LAI, chlorophyll and water content. The LAI parameter seemed to be the most affected by the object signature approach because information on intra-canopy variability of the Average Mean Angle (ALA) reduces the confounding effect between ALA and LAI (Baret and Guyot 1991).

4.2 Inversion via Iterative Optimization Techniques: Fundamentals and State-of-the-art

The solution of the canopy inverse problem using iterative optimization techniques can be stated as follows (Privette et. al. 1991, Privette et. al. 1996): Given a set of directional reflectance measurements, determine the set of independent model parameters that best fit measured directional and spectral reflectances. The best fit is determined by setting a proper cost function and a minimization procedure. Following Goel and Strebel (1983) that first defined the fitness of empirical data as “merit function”, the cost function (ϵ^2) can be expressed as following (Fang et. al. 2003):

$$\epsilon^2 = \sum_{i=1}^N \sum_{j=1}^B W_{ij} \left(R_{ij, Meas} - R_{ij, Mod}(LAI, LAD, C_{ch}, C_w, \dots) \right)^2 \quad (5)$$

Here $R_{ij, Meas}$ are the observed directional reflectances (for a given viewing geometry and sun angle), N is the number of viewing directions, B is the number of spectral bands, W_{ij} are the weights. Importantly, $R_{ij, Mod}$ are the directional and spectral reflectances simulated by the canopy reflectance model as function of the particular canopy realization. The desired canopy parameters are generally determined by applying non-linear optimization techniques. The procedure is initiated by selecting an initial guess for the canopy parameters (initial selected canopy configuration). Subsequently, the canopy RT model is run to simulate the spectral and directional reflectances corresponding to the actual canopy realization. Since the initial guess most likely generates reflectances that are different than the observed data, the cost function is not minimal. Therefore, an iterative procedure is implemented to search for a solution where ϵ^2 is at its absolute minimum. A schematic of the overall procedure is illustrated in figure (3).

Many factors influence the implementation of good inversion scheme to retrieve canopy biophysical parameters including the selection of a “good” canopy reflectance model (i.e. a model that strikes a balance between a good fit to measurements and numbers of parameters employed to fit the observed data (Jacquemoud et. al. 2000)), the selection of an appropriate cost function, the selection of the optimal number of directions and spectral bands. Moreover, regularization and sensitivity of the model parameters must be addressed.

Equation (5) represents the cost function that is commonly employed for canopy RT model inversion. However, improvements and adjustment can be done especially to address regularization. For example, a common cost function that is a variation of (5) has been employed by Meroni et. al. (2004):

$$\varepsilon^2 = \sum_{i=1}^N \sum_{j=1}^B W_{ij} \left(\frac{R_{ij,Meas} - R_{ij,Mod}(LAI, LAD, C_{ch}, C_w, \dots)}{R_{ij,Meas}} \right)^2 \quad (6)$$

The cost function in (6) is obtained by normalizing the weighted difference between modeled and observed reflectances by the absolute value of the measured reflectances. The advantage is to use a cost function that does not favor one spectral band with respect to others.

Cost functions might be modified and constrained to include a-priori knowledge and help regularize the inversion procedure (for more details see Combal et. al. 2002 and Iaquinta et. al. 1997).

The cost function minimization must be done numerically. Clearly, the choice of the best optimization algorithm depends on the mathematical properties of the function to be minimized. The most well documented, and widely employed algorithms for iterative optimization is the Quasi-Newton method (Pinty et. al. 1990, Gobron et. al. 1997, Jacquemoud et. al. 2000). The method is well received by the portion of remote sensing community working with inversion algorithms because it requires only function evaluations (no derivative approximation is computed). Other numerical techniques have been applied including the conjugate direction set method (Kuusk 1991, Liang and Strahler 1993) and the downhill simplex method (Privette et. al. 1994, Kimes et. al. 2002). Most of the available optimization algorithms have been designed to find local minima. Therefore, one of the potential limitations is dictated by the possibility for the selected algorithm to find multiple minima as opposed to a global absolute minimum. To avoid this scenario, Qui et. al. (1998) proposed running the optimization algorithms multiple times choosing random initial guesses and run a final optimization using as initial input the parameters that provide the best estimation in terms of cost function (Meroni et. al. 2004). These limitations motivated some research into applying global optimization techniques. Fang et al. (2003) applied a Genetic Algorithm Global Optimization technique to the problem of retrieving LAI from Landsat ETM+ and ground measurements. The technique scans all initial conditions and provides several possible solutions as global minimum candidates. The global character of the optimization approach avoids the inaccuracies introduced by more traditional local optimization techniques. Despite some initial interesting results in terms of LAI accuracy, more testing must be done to evaluate global performances.

Traditional inversion techniques have been historically applied to the problem of retrieving canopy biophysical parameters more than 20 years ago (e.g. Goel and Strehbel 1989). They were functional in studying the feasibility of using canopy reflectance model as effective means for canopy parameters estimation (Jacquemoud 1993, Privette et. al. 1994). Multiple studies were subsequently performed trying to improve the accuracy of the retrieval and exploring a large variety of numerical methods (Privette et. al. 1996, Qi et. al. 1998, Jacquemoud et. al. 2000, Meroni et. al. 2004). As shown above, for iterative optimization techniques, a stable and reliable inversion is not guaranteed since the searching algorithm might get stuck in a local minimum failing to reach a global minimum (Gong et. al. 1999, Kimes et. al. 2000). Moreover, traditional inversion techniques tend to be computationally expensive. Indeed, inversion must be performed

on a per-pixel basis, which involves large amount of memory storage and computational time. The latter makes the estimation of canopy biophysical parameters unfeasible for large areas (Houborg et. al. 2007). LUT and NN techniques have been recently preferred to develop operational algorithms for canopy biophysical retrieval.

4.3 Optimal Sampling

The selection of the optimal bands and directions required for accurate retrieval of canopy biophysical parameters is an important subject and it is generally investigated during the inversion algorithm development phase. Verstrate et. al. (1996) investigates some issue regarding selection of directions and spectral bands from a very theoretical point of view. Firstly, successful inversion algorithms require a number of measurements higher than the number of parameters to be retrieved (see section 4.1). Secondly, an observational sampling strategy should be devised to provide the highest density of observations where the model is most variable with respect to the independent variable considered (i.e. angle of observation, wavelength). For example, biochemical parameters can be retrieved using spectral measurements while canopy structural parameters are mostly affected by directional measurements. Thus a sampling strategy should consider what parameters are desired because if one parameter does not affect any measurement, its retrieval might be impossible (Verstrate et. al. 1996). Weiss et al (2000) conducted a study with the objective of assessing the ability to estimate canopy biophysical variables (including LAI and chlorophyll content) using several directions and visible and Near-Infrared (NIR) bands. The inversion was performed using a LUT inversion technique and the SAIL+PROSPECT model. They showed that using a limited amount of radiometric information results in poor estimation of both LAI and chlorophyll. The strong correlation between LAI and ALA (Baret and Guyot, 1991) makes the retrieval inaccurate. Subsequently, the amount of radiometric information was increased independently both directionally and spectrally. It was shown that using increasing the number of spectral bands increases the accuracy of the estimation by a factor of 2. The optimal number of bands was found to be six. Conversely, the performance of the retrieval algorithm degraded (especially for LAI) when the overall available nine bands were used. This was attributed to an increase of sensor noise. Subsequently, using the optimal six wavelengths, the best number of directions was found to be between four and seven, all located near the hot-spot or on the principal plane. Despite these initial results, the selection of an optimal sampling strategy is still an open research issue as a general criterion is not available. Meroni et. al. (2004) followed Weiss et. al. (2000) suggestion to devise the optimal bands for LAI retrieval in poplar canopies using the SAIL+PROSPECT model and iterative optimization techniques. Starting from three bands in the Red, NIR, and Shortwave Infrared (SWIR) they increased the number of bands. At each step, the inversion was performed and the optimal number of bands was selected to be the combination that reduced the Root Mean Square Error (RMSE) between retrieved LAI and measured LAI. Similarly, Schaeffl et al. (2006) after verifying that two-band radiometric information is sufficient to retrieve LAI using a hybrid model called INFORM (Atzberger, 2000) and a Neural Network approach, selected the optimal number of bands. The best combination was found by evaluating the model inversion for all possible two-band combination available in the HyMap (Schaeffl et. al. 2005) dataset. Importantly, in the case of hyperspectral data, the selection of candidate bands is performed statistically (e.g. Principal Components Analysis (PCA, Schott 1997), to avoid employing highly correlated bands that might not yield additional information and prove to increase noise.

4.4 Model Sensitivity analysis: From local to global analysis

Model sensitivity is a critical aspect of the inversion procedure. Given a set of parameters to be retrieved, the accuracy and the stability of the inversion algorithm depends on how sensitive is the model to canopy parameters variation. Ideally, the canopy RT model should have a moderate sensitivity to the parameter to be retrieved (Goel et. al. 1983). For example, if the model is not sensitive to the desired parameter, retrieval will be virtually impossible as the model will generate a very small reflectance variance for a large range of parameter variations. On the other extreme, if the model is too sensitive to the desired parameter, measurement errors will be propagate on the parameter retrieval resulting in a low accuracy of the estimation.

In a broader sense, sensitivity analysis aims at determining how the variation in a model output can be apportioned among outputs, i.e. it determines how much variation is seen by the output for changing in each input (Santner et. al. 2003). The status of knowledge in sensitivity analysis for models relevant to remote sensing and geoscience was summarized by Saltelli (1999). One decade ago, the state-of-the-art approach involved what is currently known as “local sensitivity analysis” where the derivative of the model output with respect to the desired parameter (in this context they are called sensitivity indices) is computed around a fixed point in the parameter space. For example, Iaquinata et al (1997) computed the derivative of the albedo with respect to LAI as function of the single scattering albedo and soil reflectance, determining the regions where LAI is most sensitive, i.e. visible region for bright soil and NIR region for dark soil. However, very little progresses have been achieved in the past 10 years. Bacour et. al. (2002) discusses ideas involving the design of experiment but does not compute sensitivity indices. Boyer et. al. (2003) computes the sensitivity indices for canopy RT models but he employs a large numbers of model runs. In the past few months, Morris et. al. (2008) finally addressed the problem of “global sensitivity analysis” in canopy RT models. Global sensitivity analysis refers to a statistical technique which analyzes how the model output changes when all inputs vary continuously. A global sensitivity analysis involving the explicit computation of main effects (i.e. contribution of each canopy input parameter to the computed reflectance) and sensitivity indices (i.e. expected amounts by which the uncertainty in the output is reduced if the n-th canopy input true value were known) has been performed for the LCM RT model (Ganapol et. al. 1999). The input variations are determined by a probability distribution that defines the expected values for the input parameters. Gaussian Process (GP) approximation techniques (Rasmussen and Williams, 2006) have been applied to model the hemispherical reflectance in eight MODIS bands as function of the input parameters to speed up the computation of the main effects and sensitivity indices. While their results confirmed some of the previous analysis conducted using a more constrained local sensitivity analysis (e.g. LAI is more sensitive in the NIR and chlorophyll is more sensitive in the visible, Dungan and Ganapol, 2002), the statistical framework for global sensitivity analysis developed by Morris et al. (2008) will be useful for other endeavors such as model calibration, validation and inversion. Future efforts in this field are expected to yield a more accurate quantification of model uncertainties.

4.5 Look-Up-Tables for Canopy RT Model Inversion: Fundamentals and State-of-the-art

Look-Up-Tables (LUTs) inversion is one of the most intuitive and simplest available techniques for canopy biophysical parameters estimation. It has been the basis for devising operational algorithms for LAI and fAPAR estimation using MODIS and MISR measurements (Knyazkhyn

1998a,b, Weiss et. al. 2000). An overview of the steps required to devise a LUT retrieval algorithm is shown in figure (4). Consider again equation (1): It represents the functional relationship between the space of the canopy realizations and the space of the measurements. To discuss inversion via a LUT approach, we rephrase the inverse problem in the following way: given a specific geometric configuration (sun angle geometry, sensor view angle, wavelengths) and given the sensor measurements (spectral and directional reflectances), find, within the space of all possible canopy realizations, the canopy configurations such that a match between modeled and observed data is achieved. Assuming that the relationship in equation (1) can be effectively represented by at least a canopy RT model, the LUT method comprises four step: a) generate a table in the canopy input parameter space, i.e. the space of canopy realizations that contains the input variables of the RT model (sampling of the space); b) run the selected canopy RT model in direct mode to generate the correspondent reflectance table for all desired measurement configurations (i.e. all available directions and wavelengths). This table(s) is a static element of the algorithm and it is commonly called “Look-Up-Table”; c) Given a set of observed measurements acquired during the operational phase, for a given set bands and directions, the table is scanned to search for the element(s) that are the closest (according to a selected criterion) to the observed reflectances. The corresponding element in the canopy realization table is determined to be the solution of the problem; d) If more than one solution is found, a regularization approach is implemented to converge to one result.

Few major issues must be addressed for a successful LUT-based inversion. Firstly, a criterion for selecting the closest “distance” must be defined. This is equivalent to define a cost function in the case of traditional optimization-based inversion techniques. A typical choice for the function is the so-called relative RMSE defined as follows (Weiss et. al. 2000):

$$RRMSE = \sqrt{\frac{1}{N \times B} \sum_{i=1}^N \sum_{j=1}^B W_{ij} \left(\frac{R_{ij,Meas} - R_{ij,Mod}(LAI, LAD, C_{ch}, C_w, \dots)}{R_{ij,Meas}} \right)^2} \quad (7)$$

As in the case of equation (6), this choice does not put emphasis on any particular band. Similar choices are made for the LAI/fAPAR MODIS/MISR retrieval algorithms (Knyazkhyn 1998a,b). Sampling the canopy realizations space is another important issue. Generally, the sampling is driven by what is known about the RT model input parameters distributions. The latter is intimately connected to the regularization problem, as a-priori knowledge helps to regularize the inversion algorithm (Combal et. al. 2002, see also section 4.1). A typical approach is to randomly draw the canopy realizations from the available probability distribution laws (Combal et. al. 2002, Weiss et al. 2000, Knyazkhyn et. al. 1998a,b).

Because of modeling errors and measurement uncertainties, the functional relationship between canopy parameters and measurements is not known exactly (see section 2). This means that because of modeling errors, a single canopy realization generates a modeled reflectance falling within a domain of radius $_mod$. Conversely, because of measurement uncertainties, the “true” reflected value falls within a domain of radius $_meas$. Clearly, the statement about solving the inverse problem should be modified: The admissible solutions are all canopy realizations such that the modeled reflectance falls within the modeling error and measurement uncertain domain (Kimes et. al. 2000). Generally, both model and measurement uncertainties are assumed known and lumped in a parameter $_unc$. For a set of N bidirectional reflectance measurements, the LUT

is searched and the canopy realizations satisfying the following equation (Zhang et. al. 2000, Tan et. al. 2000):

$$\frac{1}{N} \sum_{i=1}^N \left(\frac{R_{ij, Meas} - R_{ij, Mod}(LAI, LAD, C_{ch}, C_w, \dots)}{\delta_{unc}} \right)^2 < 1 \quad (8)$$

are treated as acceptable solutions to the inverse problem. Clearly, the LUT inverse algorithm does not find one solution, but a distribution of solutions over the canopy parameter space. The mean value is generally selected to be the estimated canopy parameter of interest. Next, the most notable LUT algorithm is described in details.

4.5.1 LUT for LAI retrieval: the MODIS/MISR LAI/fAPAR operational algorithm

The most known LUT-based canopy retrieval algorithm is the MODIS/MISR LAI and fAPAR operational algorithm. Knyazkhyn et. al. (1998a,b) present the theoretical justification behind the development of the retrieval scheme. The synergistic RT model-based algorithm has been designed to use of radiometric information from MODIS (single angle, up to 7 spectral bands) and MISR (nine angles, four spectral bands). In its basic form, the algorithm is based on a three-dimensional formulation of the fundamental RT processes (see section 3) to represent the functional relationship between vegetation canopy structural and optical characteristics and the directional (angular) and spectral signature observed by MODIS/MISR. The LUT approach is applied to determine the static elements that actively interact with the algorithm, i.e. the tables containing the modeled reflectance and the corresponding canopy realizations. As with any other inversion technique, the development of MODIS/MISR LAI/fAPAR algorithms had to deal with the ill-posed nature of the inversion. Patterns for spectral reflectance and transmittance of vegetation elements (optical properties), as well as patterns in the architecture of both individual trees and the entire canopy (structural parameters) have been identified and six possible biomes (grassland, shrubs, savannas, broadleaf forest, needle forest, broadleaf crops) have been introduced. Soil and understory spectral reflectance have been parametrized as function of the biome-type, i.e. selecting a specific reflectance pattern that can vary within the selected biome. This requires a-priori knowledge of the biome-type which is acquired using previously available biome maps (Biome Classification Map, Myneni et. al. 1997). The LUTs were generated as function of the specific biome by solving a three-dimensional canopy RT model. Such models are generally computationally expensive (Running et. al. 1998) and large LUTs require a large amount of memory for storage. An LUT optimization technique has been applied to reduce the number of simulations (i.e. reduce the size of the tables) and to make the overall algorithm model independent. The radiative transfer problem is divided in several independent subproblems using the general properties of the radiative transfer equation (especially its linearity). It is found that the radiation field inside a canopy can be decomposed into the sum of a radiation field generated by a canopy with black bottom surface (“black soil problem”) and the radiation field generated by a canopy with a spectral-independent anisotropic source at the soil location (“S problem”). Using conservation of energy, i.e. the sum of canopy absorbance, reflectance and transmittance must be equal to 1, the overall reflectance has been expressed as function of the canopy absorbance and transmittance of two independent problem and the effective soil reflectance. Furthermore, eigenvalue theory (Furfaro and Ganapol, 2007a, Knyazkhyn et. al. 2005) is employed to reduce the amount of spectral information required to find the radiative field: given

1 canopy absorptance and transmittance at a specified wavelength, one can determine the variables
2 at any other wavelength (principle of spectral invariance, see Knyazikhin et. al. 2005). The
3 synergistic inversion algorithm, acquires reflectance data, user-defined model and measurements
4 uncertainties (Martonchik et. al. 1998), BCM (Myneni et. al. 1997) and try to solve retrieve
5 LAI/fAPAR according to the procedure illustrated in 4.5. A back-up algorithm based on a simple
6 relationship between LAI/fAPAR and NDVI is used in case the main LUT-based algorithm fails
7 to reach convergence, i.e. to find acceptable solutions (Knyazkhyn et. al. 1998a,b)

8 To evaluate the LAI/fAPAR MODIS/MISR algorithm performances as function of spatial
9 resolution and uncertainties in surface reflectance and BMC, the algorithm has been prototyped
10 using LASUR and LANDSAT data (MODIS, Tian et. al. 2000) and POLDER data (MISR,
11 Zhang et. al. 2000) . Further algorithm refinement and studies on the impact of input
12 uncertainties on the retrieval quality has been performed for both MODIS (Wang et. al. 2001)
13 and MISR (Hu et al. 2003). It was found that misclassification between different biomes most
14 negatively impact the quality of the retrieval. Also, by comparing retrieval between coarse and
15 fine resolution data it was shown (Tian. et. al. 2000, Wang et al. 2001) that the algorithm
16 critically depends on the spatial resolution, which requires evaluation of the data density
17 distribution function. Zhang et. al. (2000) showed that multi-angle observation decreases the LAI
18 dispersion and saturation effects, increasing the quality of LAI/FAPAR retrieval (see also Hu et
19 al. 2003).

21 Full-scale validation of the inverse algorithm for retrieval of LAI and other canopy biophysical
22 variables is a critical step in assuring that the performances of the designed algorithm are
23 adequate. Current validation activities of the MODIS/MISR algorithm products (mainly LAI) are
24 performed following three basic steps, i.e. a) sampling of LAI and ancillary data from field
25 campaign b) 2) generation of a fine resolution reference LAI map based on field data, and 3)
26 comparison of MODIS/MISR LAI product with aggregated reference LAI map (Yang et. al.
27 2006a, Hu et. al. 2007)). An international collaborative effort has been established to provide the
28 required field campaigns for ground-based data collection of LAI (ground-truth, see Tan et. al.
29 2005a, Huemmrich et. al. 2005). One of the major focuses of the investigative effort has been to
30 determine a procedure where field sampling of LAI adequately represents its spatial distribution
31 and cover the natural dynamic range within each major land cover type at the site. Generally,
32 validation site have been segmented into spectrally homogeneous patches for sampling by
33 utilizing a combination of high resolution imagery and existing land cover maps. Thus a transfer
34 function (TF) using both empirical and physically-based (RT) approaches is usually determined
35 to link field-measured LAI and high resolution radiometric data. Such TF is functional in
36 defining the high resolution LAI maps (assumed to be the reference LAI values) required for
37 comparison with MODIS LAI products. The comparison should be performed by aggregating
38 LAI reference map into MODIS-like resolution maps (Yang et. al. 2006a). Generally, a pixel-by-
39 pixel comparison between MODIS and reference LAI values is not directly executed because a)
40 the actual spatial location of the corresponding pixels in the two maps may not match because of
41 geo-location and pixel-shift errors, b) the LAI algorithm is not designed to retrieve a
42 deterministic LAI value, but instead generates a mean value from all possible solutions within a
43 specified level of input satellite data and model uncertainties (Knyazikhin et. al. 1998a,b).
44 Current practice involves the comparison at the patch scale (multi-scale approach).

45 On-going validation activities for the MODIS/MISR LAI algorithm are an integral part of
46 product assessment efforts that help diagnose algorithm deficiencies, thus resulting in

refined/revised algorithms (e.g. Hu et al. 2007) , which then are used to derive the next generation of products (e.g. improvement for Collection 4 products over the previous product generation, i.e. Collection 3). The MODIS land team has found that three key factors influence the accuracy of LAI retrievals: 1) uncertainties in input land cover (biome) data, 2) uncertainties in input surface reflectances, and 3) uncertainties from the model used to build the LUTs accompanying the algorithm (Yang et. al. 2006a, 2006b.). Uncertainty in the input land cover has been addressed by replacing the old AVHRR-based land cover map with a 1-year MODIS-based map showing success in reducing the LAI over-estimation due to misclassification of grassland with broadleaf forest (Yang et. al. 2006b). Input data quality has been addressed by performing better atmospheric correction to remove the effect of cloud and aerosol (Tan et. al. 2005b). Model uncertainty has had the adverse result of causing the failure of the main retrieval algorithm and the consequent use of the back-up algorithm (low-quality products). It was shown that snow-covered vegetation areas tend to use the back-up algorithm. While some algorithm refinement has been applied to improve the main algorithm retrieval rate, the user is always strongly recommended to check the quality flag.

In conclusion, while multiple steps have been taken to improve the quality of the next generation LAI products, the uncertainty of the RT-based retrieval has not been fully addressed yet.

4.6 Neural Network Inversion Techniques: Fundamentals and State-of-the-art

Neural Networks are intelligent algorithms capable of learning complex non-linear functional relationship between input and output (Krasnopolsky and Chevallier, 2003). They are becoming a popular and computationally efficient technique for the interpretation of remotely sensed data as well as for the estimation of canopy biophysical parameters (Smith 1993, Baret et. al. 1995, Abuelgasim et. al. 1998, Kimes 1998, Gong et. al. 1999, Danson et. al. 2003, Atzberger 2004). The overall idea behind using neural networks for canopy biophysical retrieval is to implement intelligent algorithms to learn the inverse functional relationship between input data (measurements) and output data (parameters to be retrieved). A training database is generated via canopy RT model simulation of multiple canopy realizations. The training is generally performed by iteratively changing the weights and bias of the network to minimize the distance between the network current outputs and the training outputs. Multilayer Perceptrons (MLP, Demuth and Beale, 1996) are generally employed to architect neural network algorithms (Kimes et. al. 1998). MLP are comprised of elementary computational units arranged in layers. Generally, the network acquires the BRFs (measured at defined wavelengths according to the sensor configuration) and outputs the vegetation parameters V_i . For example, a three- layer network configuration is able to represent complex non linear functions in the following form:

$$V = a_2 = f_2(W^{2,1} * f_1(W^{1,1} * BRF + b_1) + b_2) \quad (9)$$

$W^{1,1}$, $W^{2,1}$, b_1 , b_2 are the synaptic weights and bias of the hidden layer and output layers and f_1 and f_2 are the neurons activation functions (Demuth and Beale 1996). Each layer is comprised of a fixed number of neurons defined by the network architect. Each neuron transforms the sum of the weighted (and biased) signals from the previous neurons according to a given transfer function and bias. The most popular transfer functions are logarithm-sigmoid and tangent-sigmoid functions (Demuth and Beale, 1996). The weights and bias of both layers are the free

parameters that can change during the training process. Once the architecture (i.e. neurons configuration) is defined, a Mean Square Error (MSE), i.e. sum of the square errors between network outputs and training outputs, is set to be the cost function. MSE is function of the weights and bias which are changed during the learning processes via iterative optimization, e.g. backpropagation algorithm (Atkinson and Tatnall 1997). No indication is given regarding the optimal number of layers and neurons included in the network architecture and generally, the best architecture is determined by trial and error (e. g. Bacour et. al. 2006). Data are pre-processed such that inputs are scaled to fall in the range [-1, 1] and both mean and standard deviations are normalized. The training set is divided in two portions, one of which is used for proper training and the other for testing the ability of the network to generalize (i.e. the ability to learn the functional relationship outside the training points) during the learning process (Demuth and Beale, 1996). Post-processing is performed to evaluate linear regression performances and data correlation between training and NN outputs. Figure (5) shows an example of neural network design, training and validation for simultaneous LAI and chlorophyll retrieval. The designed 5-layer NN has been trained using the coupled Leaf-Canopy RT model (Ganapol et. al. 1999, see section 3.3)

Many examples of NN design, training and validation are available in the literature to show how such model-based intelligent algorithms can improve the quality of canopy biophysical retrieval. Initially, the focus was on studying the feasibility of using neural networks to invert canopy RT models. For example, Smith (1993) used a MLP to invert a multiple scatter model in the optical region and predict LAI from few spectral bands. Baret et al. (1995) used a neural network to invert the combined SAIL+PROSPECT model simulating reflectances in sugar beet canopies. Fourty and Baret (1997) used a neural network to invert a coupled model (including leaf, atmospheric, canopy and soil models) simulating reflectances in the NIR-SWIR domain (up to 3 microns). More recently, many authors designed and validated neural network inversion algorithms that became the backbone of operational algorithms. Fang and Liang (2003b) designed, implemented and validated a neural network algorithm to retrieve LAI for LANDSAT 7 ETM+ data. They trained a neural network on both canopy models and coupled canopy plus atmospheric models. The network employed band 3 and 4 as well as a combination of band 4 and NDVI. Comparisons with LAI ground measurements showed a correlation ranging from 0.8 to 0.84 (R^2). Interestingly, band 3 and 4 was recommended for top-of-the-canopy retrieval and band 4 and NDVI for top-of-the-atmosphere retrieval. Schaefl and Atzberger (2006) used a neural network to invert a hybrid canopy reflectance modeled (INFORM, Atzberger 2000). The goal was to use hyperspectral data to retrieve LAI. A simple network (one neuron in the hidden layer) acquiring two bands (NIR and mid-IR) shows good agreement with LAI ground measurements (RMSE = 0.66 and R^2 = 0.74). Bacour et al. (2006) devised a neural network approach to operationally and simultaneously retrieve four canopy biophysical parameters (LAI, fAPAR, fraction of vegetation cover (fCov), the product of chlorophyll and LAI) from MERIS images (See section 5.4 for instrument details). Eleven spectral bands plus three angles defining the observation geometry were used as NN inputs. The training set was generated using SAIL+PROSPECT model coupled with a-priori information about canopy biophysical parameters (knowledge of parameters distribution laws). The training set was also corrupted with a 4% gaussian noise (zero bias) to introduce measurement uncertainties as estimated from MERIS. The retrieved variables have been compared with other products from MODIS and MERIS as well as validated against ground measurements. Good agreement with MODIS

products is shown. However, the NN tend to estimate lower LAI in broadleaf and needle forests. This is probably due to the inability of the SAIL model to account for LAI clumping exhibited by real canopies. Good agreement with ground data is shown especially for LAI (RMSE = 0.94). Baret et. al. (2007) describes the principles behind the design of a neural network algorithm to estimate LAI, fAPAR and fraction of vegetation cover from the VEGETATION sensor (see section 5.3 for instrument details). Using SAIL+PROSPECT, three independent neural network algorithms were designed and trained, one for each parameter to be retrieved. This contrasts with the approach followed by Bacour et. al. (2006), who implemented an algorithm for simultaneous parameter retrieval. Each NN acquires four inputs (the VEGETATION bands B_2 (645 nm), B_3 (835 nm) an SWIR (1165 nm) plus the sun angle) and provided one output (either LAI, fAPAR or fCov). Theoretical performance analysis showed an improvement in LAI (RMSE = 1.1), fAPAR (RMSE = 0.08) and fCov (RMSE = 0.06). LAI poorest performances are attributed to the sensitivity of LAI for larger values.

Canopy RT model-based NN inversion techniques find application also in the field of precision agriculture. Indeed, a canopy RT model has been employed to design NNs for coffee ripeness prediction (Furfaro et. al. 2005, Furfaro et. al. 2007b). Three bands (two visible and one NIR) were acquired using a commercial camera (DuncanTech MS3100, Redlake Inc.) mounted on the high-altitude, solar-powered, Helios UAV (Herwitz et. al. 2002). The algorithm was trained on synthetic data simulated using the LCM RT model (Ganapol et. al. 1999). Since a-priori information were scarcely available, a regularization technique called “Domain Projection Technique” (DPT) was implemented to systematically project the reflectance in a region where physical solutions were available. A set of 27 neural networks was employed to find the best available solution. Correlation between predictions and yield data across all ripeness levels was 0.78 with a mean absolute error of 11% (ranging from 1% to 26%).

4.7 Performances and Comparison of LAI Operational Algorithms

A comparison between various satellite data products is required to 1) assess the performances of various LAI retrieval algorithms and 2) assess the quality of the retrieval and the effect of uncertainties. Garrigues et. al. (2008) performed a detailed study to analyze the performances of four LAI satellite products including A) ECOCLIMAP climatology (Masson et al. 2003), B) GLOBCARBON (Deng et al., 2006), C) CYCLOPES (Baret. Al. 2007) and D) MODIS Collection 4 (Knyazikhin et. al. 1998a,b, Yang et. al. 2006a,b). The CYCLOPES and MODIS products (using SPOT/VEGETATION and TERRA/MODIS observations) use the retrieval algorithms described in section 4.6 and 4.5 respectively. ECOCLIMAP is based on a global land classification surface map (15 main surface-type) derived from a combination of several land maps. The main LAI retrieval algorithm is based on an empirical relationship between Normalized Difference Vegetation Index (NDVI) and LAI, where LAI is assigned via ground measurements for each type of land class. GLOBCARBON estimates LAI from a combination of SPOT/VEGETATION and ENVISAT/AATSR sensors. Its LAI retrieval algorithm (Deng et. al. 2006), relies on land-cover specific, model-based relationships between LAI and a combination of bands (Red, NIR and SWIR). The relationships were derived using Four Scale canopy reflectance model (Chen and Leblanc, 2001). The LAI products of the proposed algorithms were compared on a common spatial (1/11.2 deg sampling) and temporal (one month) resolution using the BELMANIP network sites (Baret et. al. 2006). The intercomparison and uncertainty analysis were performed using 56 LAI reference maps obtained by ground measurements. Table 1 shows

the main features associated to the LAI products. Summarizing the results obtained by the Garrigues et. al. (2008) comparative study, the authors found that overall, CYCLOPES LAI estimates tend to be the closest to the reference map despite the inability of the algorithm to faithfully represent LAI clumping. Indeed, CYCLOPE retrieval algorithm comprises a neural network trained on SAIL (Baret et. al. 2007) which computes directional reflectances assuming homogeneous canopies. The inability of SAIL to represent LAI clumping is reflected on poor CYCLOPE LAI dynamic range (0-4) whereas all other algorithms seem to have adequate LAI range. CYCLOPES and MODIS perform very well on crops and grassland, while MODIS overestimates LAI over deciduous broadleaf and needle forests. All algorithms retrieve a continuous and smooth LAI on a global scale. However, on a continental scale, ECOCLIMAP poorly describes the surface heterogeneity and GLOBCARBON shows random LAI variations disconnected from the surface properties. CYCLOPES, MODIS and GLOBCARBON show consistent temporal profiles and realistic seasonal and inter-annual variations.

5. Satellite-based Sensors for Vegetation Monitoring

Since the launch of Landsat Multispectral Scanner System (MSS, 1972), many satellite-based sensors have been placed in orbit around Earth to collect the spectral and directional electromagnetic radiation reflected by its surface. Here, we provide an overview of the most important satellite-based sensors (the US-built MODIS and MISR, and the EU-built VEGETATION and MERIS) that have been employed to observe Earth vegetation and retrieve canopy biophysical parameters.

5.1 MODIS

One of the most important instruments ever deployed to observe vegetation is the Moderate Resolution Imaging Spectrometer or MODIS (Justice et. al. 1998, see figure 6, A,C). Two MODIS instruments are currently flying on-board Terra (Earth Observation System AM) and Aqua (Earth Observation System PM) spacecrafts. Both missions have been designed to acquire data that critically contribute to advance our understanding of Earth global dynamics and processes occurring on the surface, including land and oceans. Both Terra and Aqua have been placed on a 705-km near-polar, sun-synchronous orbit with descending node at 10:30am (equatorial crossing North-South time for Terra) or ascending node at 1:30pm (equatorial crossing South-North time for Aqua). The MODIS instrument observes Earth's surface using 36 bands ranging from 0.4 microns to 14.4 microns each with a very high radiometric sensitivity (12 bits). MODIS has a swath size of 2330 km cross-track and 10 km along track (nadir). The spatial resolution is function of the band: two bands are imaged at 250 m, five bands at 500 m and 29 bands at 1km spatial resolution. The main subcomponent of MODIS instrument is the Scan Mirror Assembly. The system uses a continuously rotating double-sided mirror scanning from -55 to +55 degrees. The two-mirror, off-axis telescope has been designed to direct the captured radiant energy to four refractive assemblies designed to cover visible, NIR, SWIR and Long Wave Infrared (LWIR). The instrument weights 228.7 kg, has a single-orbit average power of 162.5 W and transmits data at 10.6 Mbps rate.

5.2 MISR

The Multi-angle SpectroRadiometer (MISR, Diner et. al. 1998, see figure 6, B, D) was launched on board the first NASA Earth Observing System spacecraft (EOS-AM1), during late 1998. Built by Jet Propulsion Laboratory (JPL), MISR was designed to study Earth's ecology and climate via multi-angle observations during the daylight. MISR provides global, radiometric, calibrated, georectified and spatially coregistered imagery sampling the reflected electromagnetic energy in nine viewing angles and four visible and NIR bands. The rationale behind building a MISR-like instrument was to gather simultaneous directional measurements of the radiant field to characterize the anisotropy of the observed surface and determine its physical properties. The system is comprised of nine cameras, four of which point forward (A_f, B_f, C_f, D_f), other four point aftward (A_a, B_a, C_a, D_a) and the ninth points in the nadir direction (A_n). The inclination angles with respect to the local vertical both in forward and aftward directions are 26.1, 45.6, 60, and 70.5 degrees. Such configuration enables the design of canopy RT-based algorithms that exploit directional information (e.g. MISR LAI/fAPAR retrieval algorithm, Knyazkhin et. al. 1998b). The selection of the angle for the A-type cameras is driven by their use in stereoscopic image mapping of clouds (26.1 degrees provide the best base-to-height ratio for stereo work). The B cameras are set at 45.6 degrees to provide the highest sensitivity to aerosol properties. The C cameras angle (60 degree) minimizes the reflectance variation among different types of clouds. The D cameras provide the inclination angle (70.5 degrees) as high as possible compatible with practical limitations. The nine cameras observe the surface in a pushbroom mode with a swath width of about 360 km and length of 20,000 km/orbit. Using 1504-per-line CCDs, the cross-track spatial resolution of off-nadir cameras is about 275 km while the nadir camera is set to have a 250m spatial resolution. This is called local mode (Earth observed at maximum resolution). Global mode implies observation at lower spatial resolution, averaging 4x4 or 2x2 pixels. The line repeating time is calibrated for the 705 km orbit and it is 40.8 milliseconds. Each of the MISR cameras, has four different colours, i.e. blues (446 nm), green (558 nm), red (672 nm) and NIR (867 nm).

MISR has a high sensitivity to deal with a wide range of scene reflectance ranging from dark (0.02% for aerosol over water) and high reflectance contrast scenes (100% for clouds over oceans) without changing gain. A Signal-to-Noise Ratio (SNR) greater than 100 is required in dark scenes, while SNR greater than 300 is necessary to observe ocean colors. The instrument has been built with a SNR greater than 700 at full signal and highest spatial resolution. The absolute radiometric uncertainty was set to be less than 3%.

Overall, with a mass of 149 kg and an average power of 83 W (113 W peak), MISR is very capable instruments that allow characterization of the surface and atmosphere, including vegetation, land, oceans, clouds and aerosol.

5.3 VEGETATION

The VEGETATION instruments family (Baret et. al. 2007) has been designed to monitor Earth surface at a near daily global coverage with a spatial resolution of about 1.1km. The European Space Agency (ESA) launched the first VEGETATION sensor aboard the Systeme Probatoire de l'Observation de la Terre (SPOT4) satellite in 1998 and VEGETATION 2 on board SPOT5 in 2002 as part of an 11 years continuous observation program. Both instruments were placed on

Sunsynchronous polar orbit with a period of 26 days and 10:30am as equatorial crossing time. The orbit and the instrument angular aperture ensure a large swath (about 2200 km) for global coverage of the equatorial region (100% every two days from -35 to +35 deg of latitude). The instrument is a wide field of view sensor designed to employ a linear array of 1728 CCDs (working in pushbroom mode) for an overall spatial resolution of 1.15km/pixel. The system is designed to acquire spectral information in four bands. The two visible bands are designed to cover the range 440-475 nm for the Blue band and the range 620-690 nm for the Red band. The NIR band spans from 780 to 880nm and the SWIR band covers the range 1570-1690nm. The Red, NIR and SWIR bands have been selected to provide information about LAI, the absorption by the green component and water content. The Blue band adds some information about atmospheric components for atmospheric correction. A new optical design has been implemented that allow all images to be registered with an accuracy of 300 m. VEGETATION exhibits a radiometric resolution of 0.003 and an absolute radiometric uncertainty less than 5 %.

5.4 MERIS

The Medium Resolution Imaging Spectrometer (MERIS, e.g. Bacour et. al. 2006) is one of the European Space Agency (ESA)- built instruments on-board the ENVISAT Earth Observing Satellite (ENVISAT EOS). MERIS was initially conceived for ocean color observations but it has been recently used to observe land and atmospheric processes including vegetation (Baret et. al. 2006). Overall, the instrument is a very flexible tool extremely functional in helping determining features of Earth surface and atmosphere. Launched with ENVISAT EOS in 2002, MERIS is a pushbroom imaging spectrometer collecting visible and NIR radiation reflected by Earth surface during daytime. Its swath is 1150 km and is commonly divided in 5 segments each covered by a dedicated camera with slightly overlapping field of view. Each camera is equipped with a two dimensional CCD array to provide both spectral and spatial information. To enable both global coverage and detailed regional imaging, two spatial resolutions have been established. Full resolution data are acquired at 300 m spatial resolution (mainly land and coastal zones). Reduced resolution data are acquired at 1.2 km spatial resolution. Global Earth coverage is achieved within a three-day period. MERIS is capable of acquiring data in 15 bands spanning the 390-1040nm range. Importantly, the instrument is adaptive as it can change band position, width and gain anytime during the nominal mission. The spectral width can be varied from 1.3 nm to 30nm. Thus, MERIS enables multi-disciplinary data acquisitions and investigations. The radiometric performances vary depending on the observed surface or portion of the atmosphere. The instrument was designed to have a large dynamic range to image both low-level signal (dark) scenes and high-level signal (bright) scenes. MERIS most stringent radiometric requirement was to have enough radiometric sensitivity to discriminate between 30 classes of pigment concentrations commonly found in open ocean waters and spanning three orders of magnitude. The latter required designing the instrument with radiometric sensitivity of 2×10^{-4} noise equivalent spectral reflectance.

6. Conclusions

In this paper, an overview/review of the methods, techniques and problems encountered in retrieving canopy biophysical parameters is presented. While a variety of methods is available in the literature, we emphasized the connection between canopy RT models, their inversion and the development of operational algorithms designed to retrieve canopy parameters. The most

important operational algorithms are based on inverting physically-based models that are firmly grounded on canopy RT theory. We reviewed the fundamentals of photon transport in turbid media and show the basic equations that must be solved to successfully simulate spectral and directional reflectance as function of canopy structural and biochemical parameters. After introducing few well-known one- and three-dimensional RT models commonly used to predict the radiative regime within and top of the canopy, the inversion problem was discussed. Traditional iterative optimization, Look-Up-Tables and Neural Networks techniques were discussed from both a theoretical and practical point of view. Some of the common issues, such as sensitivity analysis, optimal sampling and regularization were highlighted and the methods currently available in the literature to mitigate the problems were reviewed. With special emphasis on the LAI retrieval process, the most common operational algorithms for estimating canopy biophysical parameters were analyzed and compared. Finally, an overview of the functional characteristic of two US-built satellite-based sensors (MODIS and MISR) and two EU-built sensors (VEGETATION and MERIS) was presented.

As the science and technology of remote sensing progresses, some of the RT modeling and methods that represent the current state-of-the-art will be considered obsolete. However, our understanding of the physics underlying the interaction between photons and vegetation as well as our understanding of the basic nature of the inversion process, will guide scientists and engineers toward the quest for novel methodologies yielding more accurate, efficient and stable retrieval algorithms.

7. Acknowledgements

This work was partially supported by NASA Applied Information System Research (AISR) program (NRA: NNX07AV69G).

8. References

- Abuelgasim, A. A., Gopal, S., & Strahler, A. H. (1998). Forward and inverse modelling of canopy directional reflectance using a neural network. *International Journal of Remote Sensing*, 19(3), 453_471.
- Atkinson, P. M., & Tatnall, A. R. L. (1997). Neural networks in remote sensing. *International Journal of Remote Sensing*, 18(4), 699_709.
- Atzberger, C. (2000). Development of an invertible forest reflectance model: The INFOR-Model. In Buchroithner (Ed.), *A decade of trans-European remote sensing cooperation. Proceedings of the 20th EARSeL Symposium Dresden, Germany, 14– 16. June 2000* (pp. 39–44).
- Atzberger, C., Schlerf, M., Jarmer, T., Kftz, B., & Werner, W. (2003). Retrieval of wheat canopy attributes from hyperspectral HyMAP data and physically based canopy reflectance models. In M. Habermeyer, A. Muller, & S. Holzwarth (Eds.), *Proc. 3rd EARSeL Workshop on Imaging Spectroscopy, Herrsching, 13–16 May 2003*, (pp. 473–482). ISBN 2-908885-26-3.

- 1 Atzberger, C. (2004). Object-based retrieval of biophysical canopy variables using artificial
2 neural nets and radiative transfer models. *Remote Sensing of Environment*, 93, 53_67.
- 3
- 4 Baret, F., & Guyot, G. (1991). Potentials and limits of vegetation indices for LAI and APAR
5 assessment. *Remote Sensing of Environment*, 35, 161– 173.
- 6
- 7 Baret, F., Morisette, J., Fernandes, R., Champeaux, J. -L., Myneni, R. B., Chen, J., et al. (2006).
8 Evaluation of the representativeness of networks of sites for the global validation and inter-
9 comparison of land biophysical products. Proposition of the CEOS-BELMANIP. *IEEE*
10 *Transactions on Geoscience and Remote Sensing*, 44(7), 1794_1803.
- 11
- 12 Baret, F., et al. (2007), LAI, fAPAR and fCover CYCLOPES global products derived from
13 VEGETATION: Part 1: Principles of the algorithm, *Remote Sens. Environ.*, 110, 275– 286.
- 14
- 15 Bell, G. I., and S. Glasstone (1970) *Nuclear Reactor Theory*. New York: Van Nostrand Reinhold,
- 16
- 17 Bicheron, P., & Leroy, M. (1999). A method of biophysical parameter retrieval at global scale by
18 inversion of a vegetation reflectance model. *Remote Sensing of Environment*, 67, 251–266.
- 19
- 20 Buermann, W., Wang, Y., Dong, J., Zhou, L., Zeng, X., Dickinson, R. E., et al. (2002). Analysis
21 of multiyear global vegetation index data set. *Journal of Geophysical Research*, 107(D22), 4646.
- 22
- 23 Bunnik, N. (1978). *The multispectral reflectance of shortwave radiation by agricultural*
24 *crops in relation with their morphological and optical properties*. Pudoc Publications,
25 Wageningen, The Netherlands.
- 26
- 27 Case, K.M., Zweifel, P.F., (1967) *Linear transport theory*, Addison-Wesley, Reading.
- 28
- 29 Chandrasekhar, S. 1950, *Radiative Transfer*. New York: Dover, 1950.
- 30
- 31 Chen, J. M., and S. Leblanc (2001), Multiple-scattering scheme useful for hyperspectral
32 geometrical optical modelling, *IEEE Trans. Geosci. Remote Sens.*, 39(5), 1061– 1071.
- 33
- 34 Combal, B., Baret, F., Weiss, M., Trubuil, A., Mace', D., Pragne're, A., et al.(2002). Retrieval of
35 canopy biophysical variables from bidirectional reflectance using prior information to solve the
36 ill-posed inverse problem. *Remote Sensing of Environment*, 84, 1– 15.
- 37
- 38 Curran, P. J. (1994). Imaging spectrometry. *Progress in Physical Geography*, 18(2), 247– 266.
- 39
- 40 Danson, F. M., Rowland, C. S., & Baret, F. (2003). Training a neural network with a canopy
41 reflectance model to estimate crop leaf area index. *International Journal of Remote Sensing*,
42 24(23), 4891_4905.
- 43
- 44 DeWit, C. (1965). *Photosynthesis of leaf canopies*. Technical Report Agric. Res. Report 663,
45 Pudoc Publ., Wageningen, The Netherlands.
- 46

- 1 Dungan , J. and Ganapol B., 2002, Sources of uncertainty in the prediction of LAI/fPAR
- 2 fromMODIS, in Proc. AGU Fall Meeting, San Francisco, CA, 2002.
- 3
- 4 Fang, H., Liang, S., & Kuusk, A. (2003). Retrieving leaf area index using a genetic algorithm
- 5 with a canopy radiative transfer model. Remote Sensing of Environment, 85, 257–270.
- 6
- 7 Furfaro, R., B. D. Ganapol, L. F. Johnson, and S. R. Herwitz. 2005. Model-based neural network
- 8 algorithm for coffee ripeness prediction using Helios UAV aerial images. In Proc. SPIE Int’l.
- 9 Symp. Remote Sensing. Bellingham, Wash.: Int’l. Soc. Optical Engineering.
- 10
- 11 Furfaro, R. and Ganapol, B., 2007b, Spectral Theory for Photon Transport in Dense Vegetation
- 12 Media: Caseology for the Canopy Equation, Transp. Theory Stat. Phys. 36, 107-135, 2007.
- 13
- 14 Furfaro, R., Ganapol, B. D., Johnson, L. F., and Herwitz, S. R., 2007b, Coffee Ripeness
- 15 Monitoring with Airborne Imagery: Neural Network Algorithm for Field Ripeness Evaluation
- 16 Using Reduced a priori Information., Appl. Eng. Ag. 23:379-387, 2007.
- 17
- 18 Ganapol, B. D., and Myneni, R. B. (1992), The FN for the one-angle radiative transfer equation
- 19 applied to plant cally, the dependence of canopy reflectance on leaf canopies. Remote Sens.
- 20 Environ. 39:213–231.
- 21
- 22 Ganapol, B.D., L.F. Johnson,P.D. Hammer, C.A. Hlavka, D.L. Peterson (1998) LEAFMOD: A
- 23 new Within-Leaf Radiative Transfer Model. Remote Sensing of Environment, 63(2): 182-193.
- 24
- 25 Ganapol, B.D., L.F. Johnson, C.A. Hlavka, D.L. Peterson, and B. Bond (1999) LCM2: A
- 26 coupled leaf/canopy radiative transfer model, Remote Sensing of Environment, 70:153-166.
- 27
- 28 Ganapol, B., D., (2006), Vegetation Canopy Reflectance Modeling with Turbid Medium
- 29 Radiative Transfer, Lecture Notes in Computational Science and Engineering, Computational
- 30 Methods in Transport, Springer Berlin 2006
- 31
- 32 Gastellu-Etchegorry, J. P., Demarez, V., Pinel, V., & Zagolski, F. (1996). Modeling radiative
- 33 transfer in heterogeneous 3-D vegetation canopies. Remote Sensing of Environment, 58,
- 34 131_156.
- 35
- 36 Gastellu-Etchegorry, J. P., Martin, E., & Gascon, F. (2004). DART: a 3D model for simulating
- 37 satellite images and studying surface radiation budget. International Journal of Remote Sensing,
- 38 25, 73_96.
- 39
- 40 Gobron, N., Pinty, B., & Verstraete, M. M. (1997). Theoretical limits to the estimation of the leaf
- 41 area index on the basis of visible and nearinfrared, remote sensing data. IEEE Transactions on
- 42 Geoscience and Remote Sensing, 35, 1438– 1445.
- 43
- 44 Goel, N. S., & Strebel, D. E. (1983). Inversion of vegetation canopy reflectance models for
- 45 estimating agronomic variables: I. Problem definition and initial results using the Suits’ model.
- 46 Remote Sensing of Environment, 36, 73– 104.

- 1
- 2 Goel, N. S., & Thompson, R. L. (1984). Inversion of vegetation canopy reflectance models for
- 3 estimating agronomic variables: IV. Total inversion of the SAIL model. *Remote Sensing of*
- 4 *Environment*, 15,237–253.
- 5
- 6 Goel, N., 1987, Models of vegetation canopy reflectance and their use in estimation of
- 7 biophysical parameters from reflectance data. *Remote Sensing Reviews*, **3**, 1± 212.
- 8
- 9 Gong, P., D. Wang, and S. Liang, (1999), Inverting a canopy reflectance model using an
- 10 artificial neural network. *International Journal of Remote Sensing*. 20(1):111-122.
- 11
- 12 Govaerts, Y., & Verstraete, M. M. (1998). Raytran: A Monte Carlo ray tracing model to compute
- 13 light scattering in three-dimensional heterogeneous media. *IEEE Transactions on Geoscience*
- 14 *and Remote Sensing*, 36, 493_505.
- 15
- 16 Garrigues, S., Lacaze, F., Baret,, F., Morisette, J., T., Weiss, M., Nickeson, J., E., Fernandes,
- 17 R., Plummer, S., Shabanov, N., V., Myneni, R., B., Knyazikhin, Y., Yang, W., (2008),
- 18 Validation and intercomparison of global Leaf Area Index products derived from remote sensing
- 19 data. *JOURNAL OF GEOPHYSICAL RESEARCH*, VOL. 113, G02028,
- 20 doi:10.1029/2007JG000635
- 21
- 22 Habermann, P., (2003), *Applied Partial Differential Equations: With Fourier Series and*
- 23 *Boundary Value Problems*, Prentice Hall
- 24
- 25 Herwitz, S. R., L. F. Johnson, J. Arvesen, R. G. Higgins, J. G. Leung, and S. L. Dunagan. 2002.
- 26 Precision agriculture as a commercial application for solar powered unmanned aerial vehicles. In
- 27 *Proc. 1st AIAA UAV Conf. Portsmouth, Va.:American Inst. Aeronautics and Astronautics*.
- 28
- 29 Houborg, R., Soegaard, H., & Boegh, E. (2007). Combining vegetation index and model
- 30 inversion methods for the extraction of key vegetation biophysical parameters using Terra and
- 31 Aqua MODIS reflectance data. *Remote Sensing of Environment*, 106(1), 39_58.
- 32
- 33 Hu, J., Tan, B., Shabanov, N., Crean, K. A., Martonchik, J. V., Diner, D. J., et al. (2003).
- 34 Performance of the MISR LAI and FPAR algorithm: A case study in Africa. *Remote Sensing of*
- 35 *Environment*, 88, 324_340.
- 36
- 37 Huang, D., Knyazikhin, Y., Wang, W., Deering D., W., Stenberg P., Shabanov N., Tan B.,
- 38 Myneni, M., (2008), Stochastic transport theory for investigating the three-dimensional canopy
- 39 structure from space measurement, *Remote Sensing of Environ.*, 112:35–50, 2008.
- 40
- 41 Huemmrich K. L., J. L. Privette, M. Mukelabai, R. B. Myneni, and Y. Knyazikhin, “Time-series
- 42 validation of MODIS land biophysical products in a Kalahari woodland, africa” *Int. J. Remote*
- 43 *Sens.*, vol. 26, pp. 4381–4398, 2005.
- 44
- 45 Jacquemoud, S., & Baret, F. (1990). Prospect: A model of leaf optical properties spectra. *Remote*
- 46 *Sensing of Environment*, 34(2), 75_91.

- 1
- 2 Jacquemoud, S. (1993). Inversion of the PROSPECT + SAIL canopy reflectance models from
- 3 AVIRIS equivalent spectra: theoretical study. *Remote Sensing of Environment*, 44, 281–292.
- 4
- 5 Jacquemoud, S., Bacour, C., Poilve, H., & Frangi, J. P. (2000). Comparison of four radiative
- 6 transfer models to simulate plant canopies reflectance direct and inverse mode. *Remote Sensing*
- 7 *of Environment*, 74, 471–481.
- 8
- 9 Justice, C. O., et al., 1998. The moderate resolution imaging spectroradiometer (MODIS): Land
- 10 remote sensing for global change research. *IEEE Trans. Geosc. Remote Sens.*, 36:1228-1249
- 11
- 12 Kallel, A., Verhoef, W., Le Hégarat-Masclé, S., Ottlé, L. Hubert-Moy, C., (2008), Canopy
- 13 Bidirectional Reflectance calculation based on Adding method and SAIL formalism:
- 14 AddingS/AddingSD, *Remote Sensing of Environment*, 112,9, 3639 – 3655.
- 15
- 16 Kimes, D. S., Nelson, R., Manry, M. T., & Fung, A. K. (1998). Attributes of neural networks for
- 17 extracting continuous vegetation variables from optical and radar measurements. *International*
- 18 *Journal of Remote Sensing*, 19(14), 2639_2663.
- 19
- 20 Kimes, D. S., Knyazikhin, Y., Privette, J. L., Abuelgasim, A. A., & Gao, F. (2000). Inversion
- 21 methods for physically-based models. *Remote SensingReviews*, 18(2–4), 381_439.
- 22
- 23 Kimes, D., Gastellu-Etchegorry, J., Estevec, P.,(2002), Recovery of forest canopy characteristics
- 24 through inversion of a complex 3D model, *Remote Sensing of Environment* 79, 320– 328
- 25
- 26 Knyazikhin, Y., J. Kranigk, G. Miessen, O. Panfyorov, N. Vygodskaya, and G. Gravenhorst
- 27 (1996). Modelling Three-Dimensional Distribution of Photosynthetically Active Radiation in
- 28 Sloping Coniferous Stands. *Biomass and Bioenergy*, **11**, 189–200.
- 29
- 30 Knyazikhin, Y., Martonchik, J., Myneni, R., Diner, D., & Running, S. (1998a). Synergistic
- 31 algorithm for estimating vegetation canopy leaf area index and fraction of absorbed
- 32 photosynthetically active radiation from MODIS and MISR data. *Journal of Geophysical*
- 33 *Research*, 103 (D24), 32257– 32276.
- 34
- 35 Knyazikhin, Y., Martonchik, J., Diner, D., Myneni, R., Verstraete, M., Pinty, B., & Gobron, N.
- 36 (1998b). Estimation of vegetation canopy leaf area index and fraction of absorbed
- 37 photosynthetically active radiation from atmosphere-corrected MISR data. *Journal of*
- 38 *Geophysical Research*, 103(D24), 32239– 32256.
- 39
- 40 Knyazikhin, Y., Marshak, A., Myneni, R., B., (2005). Three-Dimensional Radiative Transfer in
- 41 Vegetation Canopies. In: A. Davis and A. Marshak [Eds], "Three-Dimensional Radiative
- 42 Transfer in the Cloudy Atmosphere," Springer-Verlag, ISBN-10 (3-540-23958-8), pages 617-
- 43 651.
- 44
- 45 Kuusk, A. (1985). The hot spot effect of a uniform vegetative cover. *Sovietic Journal of Remote*
- 46 *Sensing*, 3(4), 645_658.

- 1 Kuusk, A. (1991a). Determination of vegetation canopy parameters from optical measurements.
2 Remote Sensing of Environment, 37, 207_218.
- 3
- 4 Kuusk, A. (1991b). The hot spot effect in plant canopy reflectance. In R. I. Myneni & J. Ross
5 (Eds.), Photon-vegetation interactions. Applications
6
- 7 Kuusk, A. (1995). A Markov chain model of canopy reflectance. Agricultural and Forest
8 Meteorology, 76, 221– 236.
9
- 10 Kuusk, A. (2001). A two-layer canopy, reflectance model. Journal of Quantitative Spectroscopy
11 and Radiative Transfer, 71(1), 1 –9.
12
- 13 Lewis, P. (1999). Three-dimensional plant modelling for remote sensing simulation studies using
14 the botanical plant modelling system. Agronomie–Agriculture and Environment, 19, 185_210.
15
- 16 Li, X., and Strahler, A., (1992), Geometric-optical bi-directional reflectance modeling of the
17 discrete-crown vegetation canopy: effect of crown shape and mutual shadowing, IEEE
18 Transactions on Geoscience and Remote Sensing, 30, 276- 292.
19
- 20 Liang, S., & Strahler, A. H. (1993). An analytic BRDF model of canopy radiative transfer and its
21 inversion. IEEE Transaction on Geoscience and Remote Sensing, 31, 1081– 1092.
22
- 23 Martonchik, J.V., D.J. Diner, B. Pinty, M.M. Verstraete, R.B. Myneni, Y. Knyazikhin, and H.R.
24 Gordon (1998) Determination of land and ocean reflective, radiative and biophysical
25 properties using multi-angle imaging. *IEEE Trans. Geosci. Remote Sens.*, **36**: 1266-1281.
26
- 27 Masson, V., J. L. Champeaux, F. Chauvin, C. Meriguer, and R. Lacaze (2003), A global database
28 of land surface parameters at 1 km resolution in meteorological and climate models, *J. Clim.*, 16,
29 1261–1282.
30
- 31 Meroni, M., Colombo, R., & Panigada, C. (2004). Inversion of a radiative transfer model with
32 hyperspectral observations for LAI mapping in poplar plantations. Remote Sensing of
33 Environment, 92, 195– 206.
34
- 35 Myneni, R., G. Asrar, and S. Gerstl (1990). Radiative transfer in three dimensional leaf canopies.
36 *Transp. Theor. Stat. Phys.*, **19**, 205–250.
37
- 38 Myneni, R.B., Marshak, A. and Knyazihin, M. and Asrar, G., (1991a). Discrete Ordinates
39 Method for Photon Transport in Leaf Canopies. In: R.B. Myneni and J. Ross [Eds.], "Photon-
40 Vegetation Interactions: Applications in Optical Remote Sensing and Plant Physiology",
41 Springer-Verlag, pp. 45-109.
42
- 43 Myneni, R.B. (1991b) Modeling radiative transfer and photosynthesis in three-dimensional
44 vegetation canopies. *Agric. For. Meteorol.*, 55: 323-344.
45
- 46

- 1 Myneni, R.B., R.R. Nemani, and S.W. Running, Estimation of global leaf area index and
2 absorbed PAR using radiative transfer model, *IEEE Trans. Geosci. Remote Sens.*, 35, 1380-1393,
3 1997b.
- 4
- 5 North, P. (1996). Three-dimensional forest light interaction model using a Monte Carlo method.
6 *IEEE Transactions on Geoscience and Remote Sensing*, 34, 946_956.
- 7
- 8 Pinty, B. and M. Verstraete (1997). Modeling the scattering of light by vegetation
9 in optical remote sensing. *Journal of Atmospheric Science.*, 55, 137–150.
- 10
- 11 Pomraning, G. C. (1991). Linear kinetic theory and particle transport in stochastic mixtures (pp.
12 235). New Jersey: World Scientific.
- 13
- 14 Pomraning, G. C. (1996). Transport theory in discrete stochastic mixtures. *Advances in Nuclear*
15 *Science and Technology*, 24, 47_93.
- 16
- 17 Privette, J. L., Myneni, R. B., Tucker, C. J., & Emery, W. J. (1994). Invertibility of a 1-D
18 discrete ordinates canopy reflectance model. *Remote Sensing of Environment*, 48, 89–105.
- 19
- 20 Privette, J. L., Emery, W. J., & Schimel, D. S. (1996). Inversion of a vegetation reflectance
21 model with NOAA AVHRR data. *Remote Sensing of Environment*, 58(2), 187–200.
- 22
- 23 Qiu, J., Gao, W., & Lesht, B. (1998). Inverting optical reflectance to estimate surface properties
24 of vegetation canopies. *International Journal of Remote Sensing*, 19(4), 641–656.
- 25
- 26 C. Rasmussen and C. Williams, *Gaussian Processes for Machine Learning*. Cambridge, MA:
27 MIT Press, 2006.
- 28
- 29 Ross, J., editor (1981). *The Radiation Regime and Architecture of Plant Stands*. Dr. W. Junk,
30 Norwell (MA).
- 31
- 32 Running, S.W., R.B. Myneni, R. Nemani, and J. Glassy. (1996) MOD15 LAI/FPAR algorithm
33 theoretical basis document, MODIS LAI (leaf area index) and MODIS FPAR (fraction of
34 absorbed photosynthetically active radiation). *Tech. Rep.*, Sch. of For., Univ. of Mont.,
35 Missoula.
- 36
- 37 T. Santner, B. Williams, and W. Notz, *The Design and Analysis of Computer Experiments*,
38 *Springer Series in Statistics*. Berlin, Germany: Springer-Verlag, 2003.
- 39
- 40 Schlerf, M., Atzberger, C., & Hill, J. (2005). Remote sensing of forest biophysical variables
41 using HyMap imaging spectrometer data. *Remote Sensing of Environment*, 95, 177– 194.
- 42
- 43 Schlerf, M., & Atzberger, C. (2006). Inversion of a forest reflectance model to estimate structural
44 canopy variables from hyperspectral remote sensing data. *Remote Sensing of Environment*,
45 100(3), 281_294.
- 46

- Schott J.R., (1997), "Remote Sensing: The Imaging Chain Approach," Oxford University Press 1997.
- Sellers, P. J., Randall, D. A., Betts, A. K., Hall, F. G., Berry, J. A., Collatz, G. J., et al. (1992). Modeling the exchanges of energy, water, and carbon between continents and the atmosphere. *Science*, 275, 502_509.
- Shabanov, N. V., Knyazikhin, Y., Baret, F., & Myneni, R. B. (2000). Stochastic modeling of radiation regime in discontinuous vegetation canopies. *Remote Sensing of Environment*, 74(1), 125_144.
- Shabanov, N. V., Kotchenova, S., Huang, D., Yang, W., Tan, B., Knyazikhin, Y., et al. (2005). Analysis and optimization of the MODIS leaf area index algorithm retrievals over broadleaf forests. *IEEE Transactions on Geoscience and Remote Sensing*, 43(8), 1855_1865.
- Shultis, K.J., Myneni, R.B., (1989), Radiative transfer in vegetation canopy with anisotropic scattering, *J. Quant. Spectrosc. Radiat. Transfer*, Vol. 39, 115-129.
- Sidi A., (2003), *Practical Extrapolation Methods*, Cambridge University Press, New York .
- Smith, A. J., 1993, LAI inversion using a back-propagation neural network trained with a multiple scattering model. *IEEE Transactions on Geoscience and Remote Sensing*, 31, 1102- 1106.
- Suits, G. H. (1972). The calculation of the directional reflectance of a vegetative canopy. *Remote Sensing of Environment*, 2, 117_125.
- Tian, Y., Y. Zhang, Y. Knyazikhin, R. B. Myneni, J. Glassy, G. Dedieu, and S. W. Running (2000), Prototyping of MODIS LAI and FPAR algorithm with LASUR and LANDSAT data, *IEEE Trans. Geosci. Remote Sens.*, 38, 2387– 2401.
- Tan B., J. Hu, P. Zhang, D. Huang, N. V. Shabanov, M. Weiss, Y. Knyazikhin, and R. B. Myneni, 2005a, "Validation of MODIS LAI product in croplands of Alpilles, France," *J. Geophys. Res.*, vol. 110, 2005. D01107, DOI:10.1029/2004JD004860.
- Tan B., D. Huang, J. Hu, W. Yang, P. Zhang, V. N. Shabanov, Y. Knyazikhin, and R. B. Myneni, 2005b, "Assessment of the broadleaf crops Leaf Area Index product from the terra MODIS instrument," *Agric. Forest Meteorol.*, vol. 135, pp. 124–134, 2005.
- Vanderbilt, V., L. Grant, and S. Ustin (1991), Myneni, R. and J. Ross, editors, *Polarization of light by vegetation*, pages 191–228. Springer Verlag, New York, USA.
- Verhoef, W. (1984). Light scattering by leaf layers with application to canopy reflectance modelling: The sail model. *Remote Sensing of Environment*, 16, 125_141.

- 1 Verhoef, W. (1985). Earth observation modeling based on layer scattering matrices. *Remote*
2 *Sensing of Environment*, 17, 165_178.
- 3
- 4 Verhoef, W. (1998). Theory of radiative transfer models applied to optical remote sensing of
5 vegetation canopies. Ph.D. thesis, Agricultural University, Wageningen, The Netherlands.
- 6
- 7 Verhoef, W., & Bach, H. (2003). Simulation of hyperspectral and directional radiance images
8 using coupled biophysical and atmospheric radiative transfer models. *Remote Sensing of*
9 *Environment*, 87, 23_41.
- 10
- 11 Verhoef, W., & Bach, H. (2007). Coupled soil leaf canopy and atmosphere radiative transfer
12 modeling to simulate hyperspectral multi-angular surface reflectance and TOA radiance data.
13 *Remote Sensing of Environment*, 109(2), 166_182.
- 14
- 15 Wang, Y., Tian, Y., Zhang, Y., El-Saleous, N., Knyazikhin, Y., Vermote, E., et al. (2001).
16 Investigation of product accuracy as a function of input and model uncertainties: Case study with
17 seawifs and MODIS LAI/FPAR algorithm. *Remote Sensing of Environment*, 78, 299_313.
- 18
- 19 Weiss, M., Baret, F., Myneni, R. B., Pragnere, A., & Knyazikhin, Y. (2000). Investigation of a
20 model inversion technique to estimate canopy biophysical variables from spectral and directional
21 reflectance data. *Agronomie*, 20, 3– 22.
- 22
- 23 Widlowski J.L., Taberner, M., Pinty, B., Gastellu-Etchegorry, J.P., Martin, E., et al., (2007), The
24 third RAdiation transfer Model Intercomparison (RAMI) exercise, *Journal of Geophysical*
25 *Research*, 112, D09111, doi: 10.1029/2006JD007821.
- 26
- 27 Yang, W., Tan, B., Huang, D., Rautiainen, M., Shabanov, N. V., Wang, Y., Privette, J. L.,
28 Huemmrich, K. F., Fensholt, R., Sandholt, I., Weiss, M., Nemani, R. R., Knyazikhin, Y., &
29 Myneni, R. B., 2006a, MODIS leaf area index products: from validation to algorithm
30 improvement. *IEEE Trans. Geosci. Remote Sens.* VOL. 44, NO. 7, JULY 2006.
- 31
- 32 Yang, W., Huang, D., Tan, B., Stroeve, J. C., Shabanov, N. V., Knyazikhin, Y., Nemani, R. R., &
33 Myneni, R. B., 2006b. Analysis of leaf area index and fraction of PAR absorbed by vegetation
34 products from the Terra MODIS sensor: 2000-2005. *IEEE Trans. Geosci. Remote Sens.* VOL.
35 44, NO. 7, JULY 2006
- 36
- 37 Zhang, Y., Tian, Y., Knyazikhin, Y., Martonchik, J.V., Diner, D. J., Leroy, M., & Myneni, R. B.
38 (2000). Prototyping of MISR LAI and FPAR algorithm with POLDER data over Africa. *IEEE*
39 *Transactions on Geoscience and Remote Sensing*, 38, 2402– 2418.
- 40
- 41
- 42
- 43
- 44
- 45
- 46

Figures and Tables

Table1: A summary of the most commonly used operational retrieval algorithms for LAI estimation (table modeled after Garrigues et. al. 2008)

Product Name	Canopy RT model	Inversion Algorithm	LAI Clumping	Uncertainties	A-priori Info	Reference
CYCLOPE-V3.1	SAIL+PROSPECT (1-D)	NN	No	Model and measurements at 4%	Probability distribution	Baret et. al. 2007
MODIS-C4	3-D RTM (DISORD)	LUTs	Yes	Measurements at 20%	Values for six biomes	Knyazikhin et. al. 1998a
GLOBCARBON-V1	Model-based NDVI-LAI relationships	N/A	No	Not Specified	Biome-type	Deng et. al. 2006
ECOCLIMAP	Empirical	N/A	No	Not Specified	Biome-type	Masson et. al. 2003
MERIS	SAIL+PROSPECT (1-D)	NN	No	Model and measurements at 4%	Probability distributions	Bacour et. al. 2006

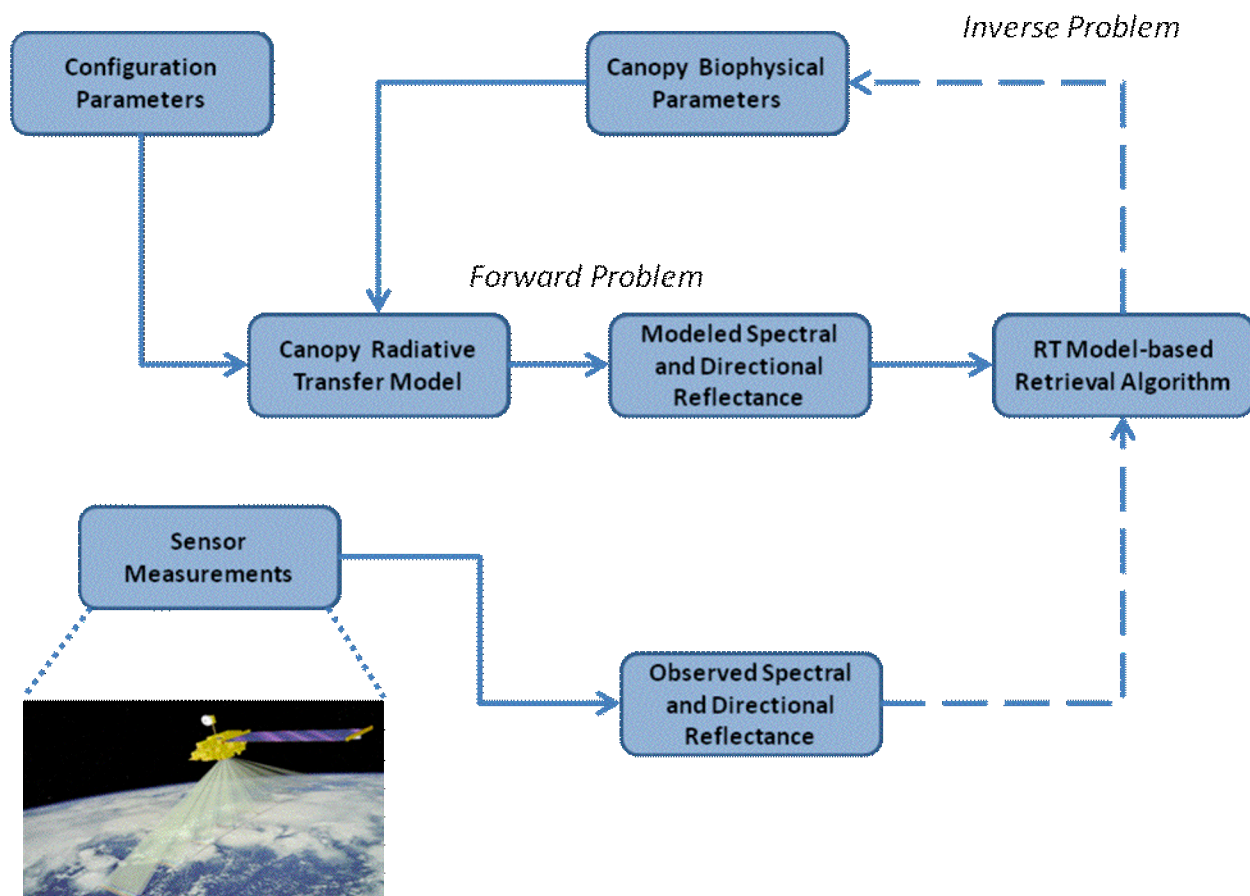


Figure 1: Diagram illustrating the canopy biophysical parameters retrieval problem in remote sensing of vegetation.

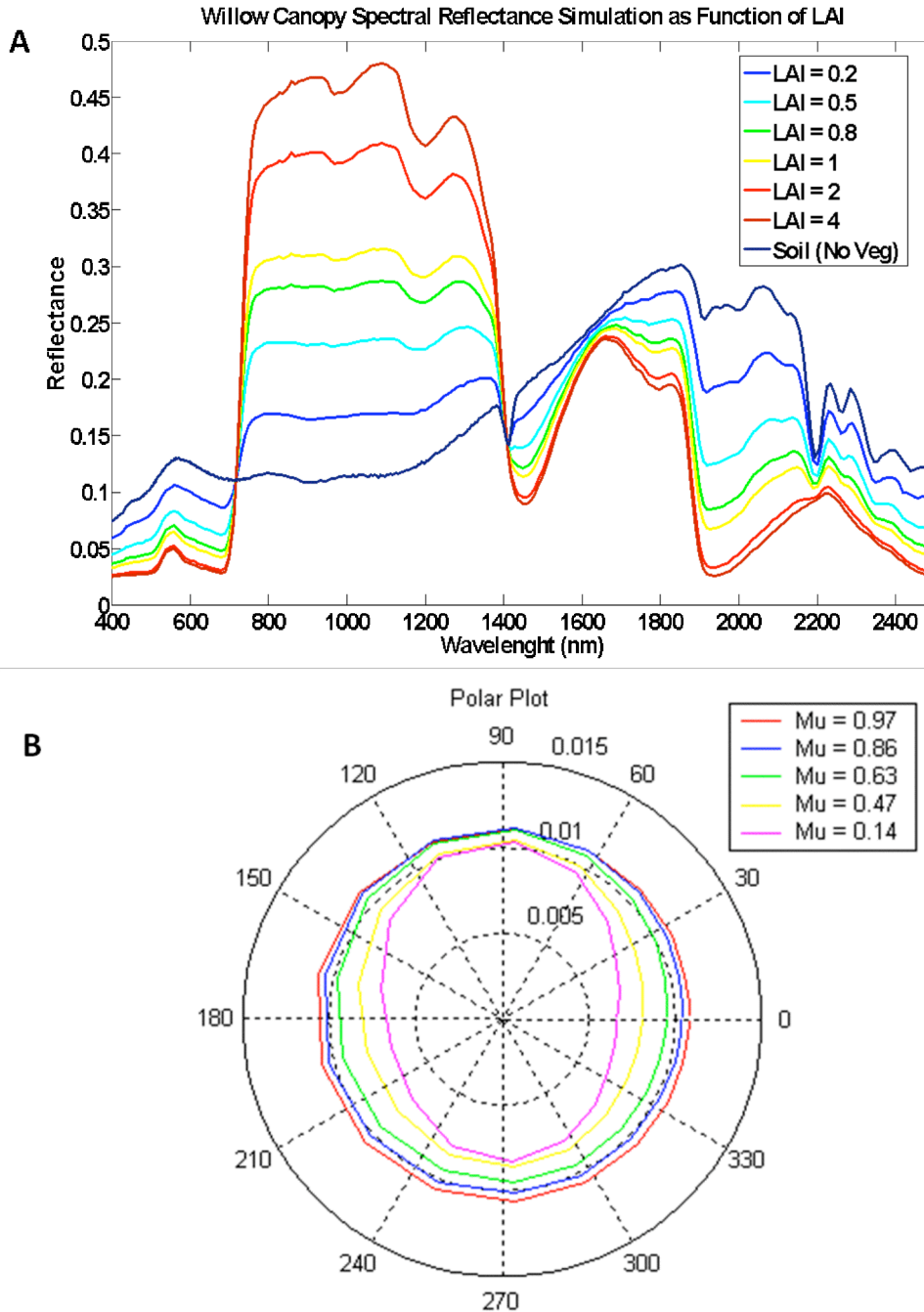


Figure 2: Examples of spectral and directional reflectance simulated by canopy RT models. A) Spectral behavior of the nadir reflectance as function of wavelength (400-2500 nm) for Willow canopies with variable LAI. The spectral behavior of the reflectance soil employed in this simulation is included. B) Polar plot for simulated directional reflectance exhibited by a maple canopy as function of view angle, (inclination and azimuth). Both simulations have been performed using the coupled Leaf-Canopy RT Model (LCM, Ganapol et. al. 1999)

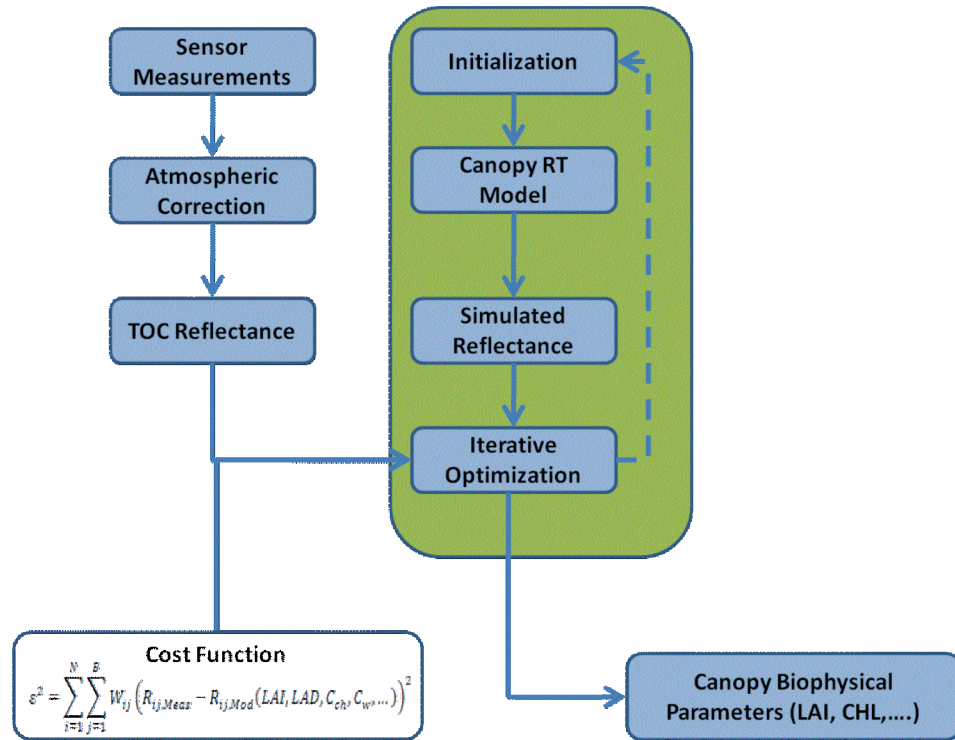


Figure 3: Conceptual scheme illustrating the steps required to retrieve canopy biophysical parameters using traditional iterative optimization techniques.

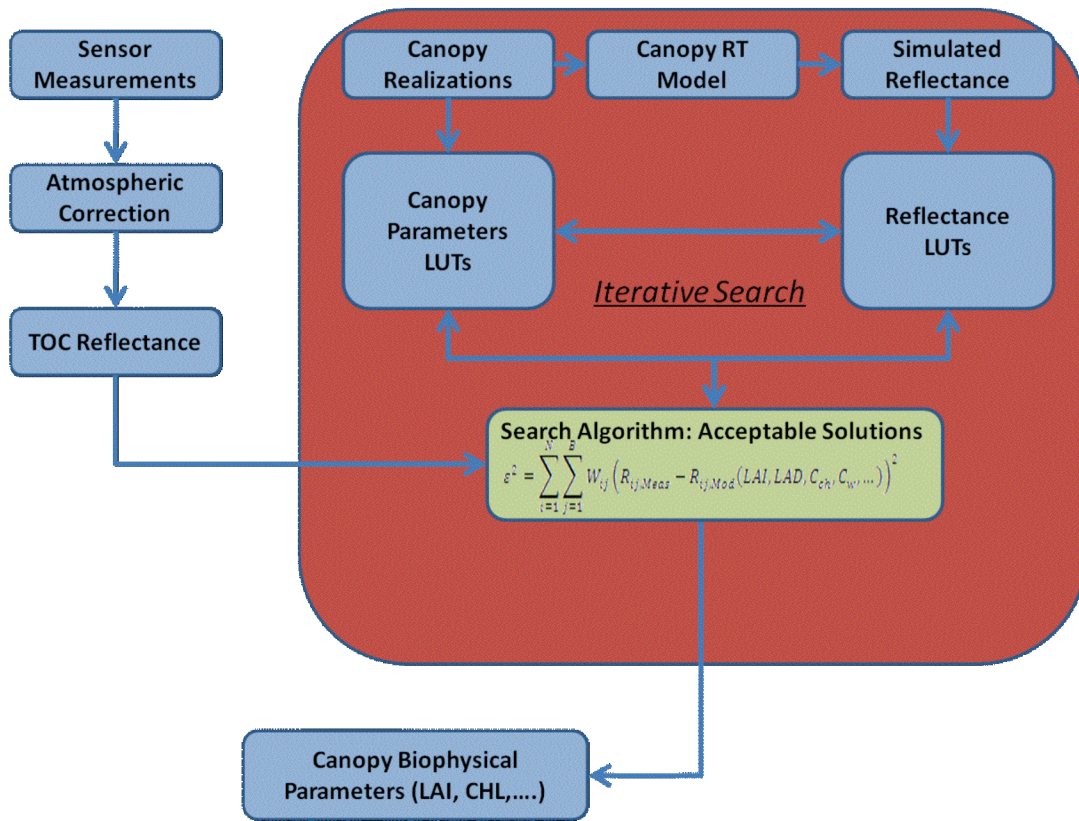


Figure 4: Conceptual scheme illustrating the step required to implement and execute the retrieval of canopy biophysical parameters using Look-Up-Tables

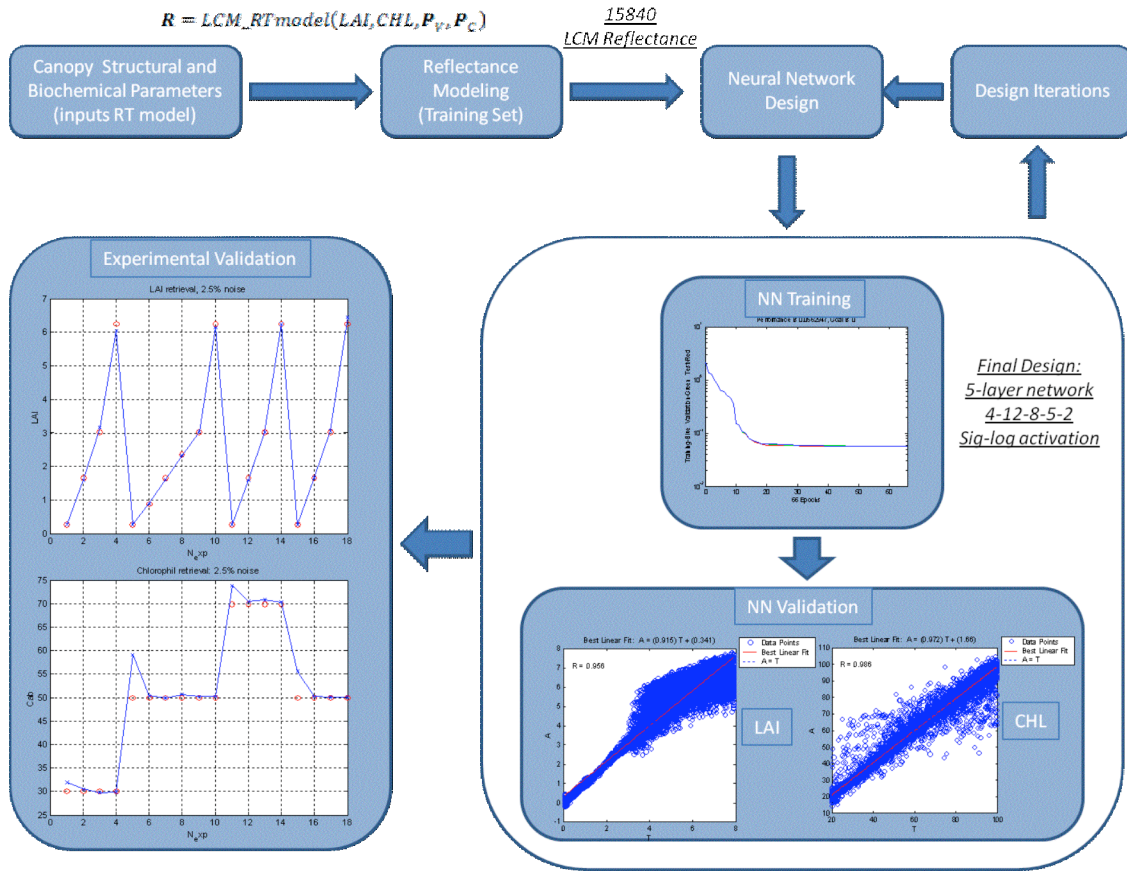
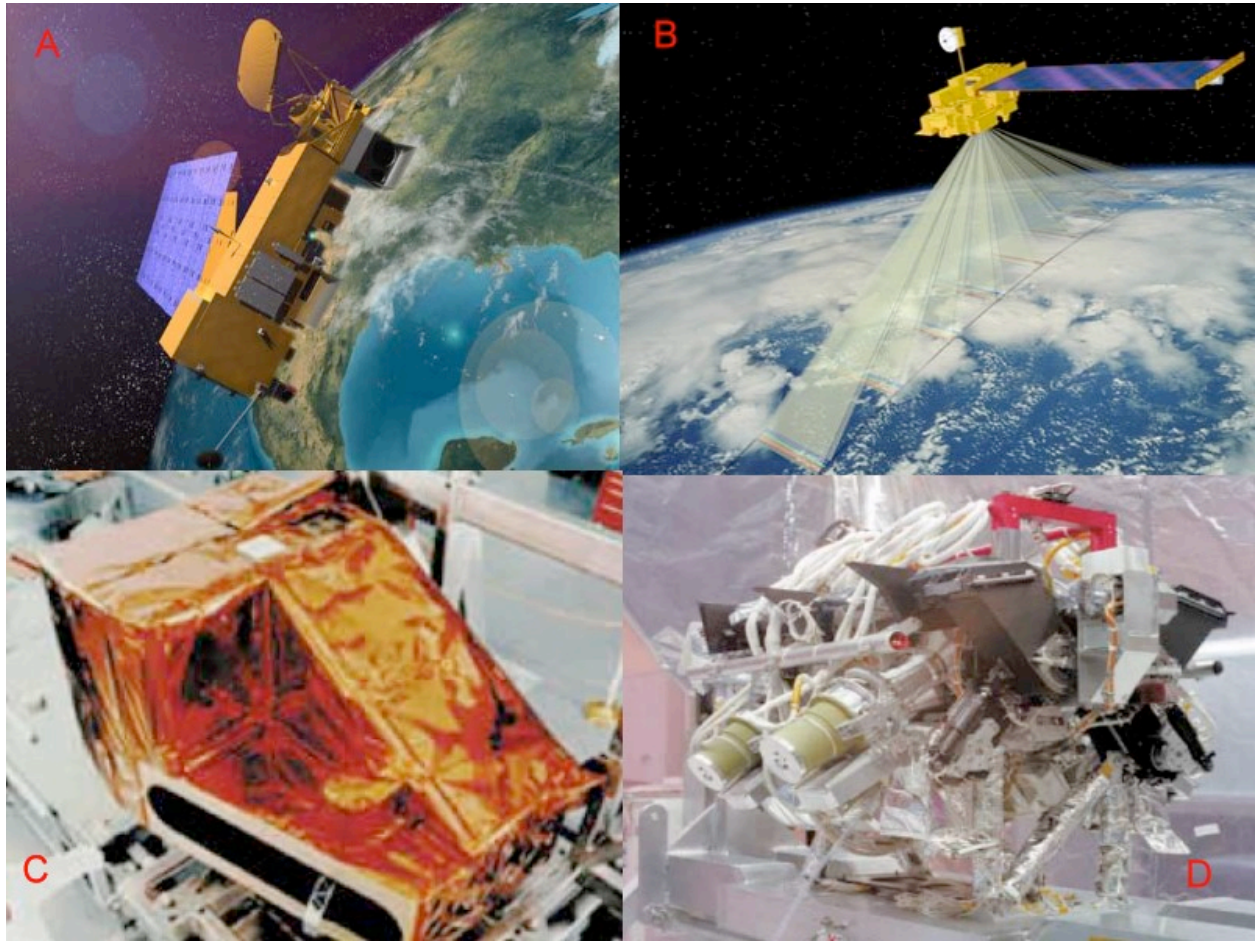


Figure 5: Canopy biophysical parameters retrieval via neural networks and canopy RT models. The presented scheme illustrates the procedure. In this example, LCM (Ganapol et. al. 1999) has been used to generate the training set for simultaneous retrieval of LAI and chlorophyll from remote sensing observation. The NN approach required a set of 15840 training points (reflectance) corrupted with a 2.5% noise and 2% relative bias. Few design iterations were necessary to select the network architecture (4-12-8-5-2). The NN acquires four inputs (nadir reflectance at 450,550,670,790 nm), uses three hidden layers to process the signals and output two parameters (LAI and chlorophyll). Training and network validation are shown at the right bottom portion of the figure. The network performances have been compared against 18 data points representing ground truth (graphs at left portion of the figure). The system is shown to retrieve LAI with RMSE equal to 0.083 and chlorophyll with RMSE equal to 2.7528. The latter is shown to be the most sensitive to noise.

1



2

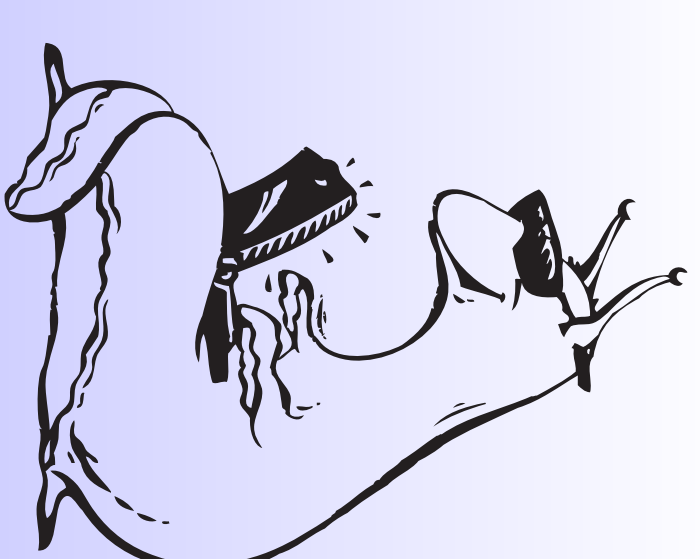
3

4 Figure 6: Two examples of instruments for remote sensing of vegetation. A and B panels are two
 5 artistic images showing the satellite AQUA (EOS PM-1) while observing Earth. C and D panels
 6 show two images of MODIS and MISR instruments, respectively. The panel B shows MISR
 7 simultaneous observation of the ground in the nine nominal directions.

BAYESIAN ESTIMATION AND SENSITIVITY ANALYSIS FOR RADIATIVE TRANSFER MODELS

Marian Farah¹, Athanasios Kottas¹, and Robin D. Morris^{2,1}, Roberto Furfaro³, Barry Ganapol³

¹Applied Math and Statistics, UC Santa Cruz, ²USRA, ³University of Arizona



Abstract

A Radiative Transfer Model (RTM) simulates the interaction of light with a medium. We are interested in RTMs that model light reflected from a vegetated region. We study the Leaf Canopy Model (LCM) RTM, which was designed to explore the feasibility of observing leaf chemistry remotely. The inputs to the LCM are leaf chemistry variables (chlorophyll, water, lignin, cellulose) and canopy structural parameters (leaf area index (LAI), leaf angle distribution (LAD), soil reflectance, sun angle), and the output is the upwelling radiation at the top of the canopy. In this work, we address the following question: which of the inputs to the RTM has the greatest impact on the computed observation? To answer this question, we employ a Bayesian Gaussian Process approximation to the LCM output using Markov Chain Monte Carlo (MCMC) simulation. Then, we analyze the “main effects” of the inputs to the LCM in terms of the sensitivity of the LCM’s output to each of the inputs. We apply this method to 7 inputs and output obtained at 667 nm and 1640 nm wavelengths, which are associated with MODIS (a key instrument aboard the Terra and Aqua satellites) spectral bands that are sensitive to vegetation.

Introduction

- The LCM was developed in order to capture the essential biophysical processes associated with the interaction between light and vegetation by combining two different radiative transfer algorithms.
- LEAFMOD simulates the radiative regime inside the single leaf, and CANNMOD combines the information coming from LEAFMOD with canopy structural parameters to compute the radiative regime within and at the top of the canopy.

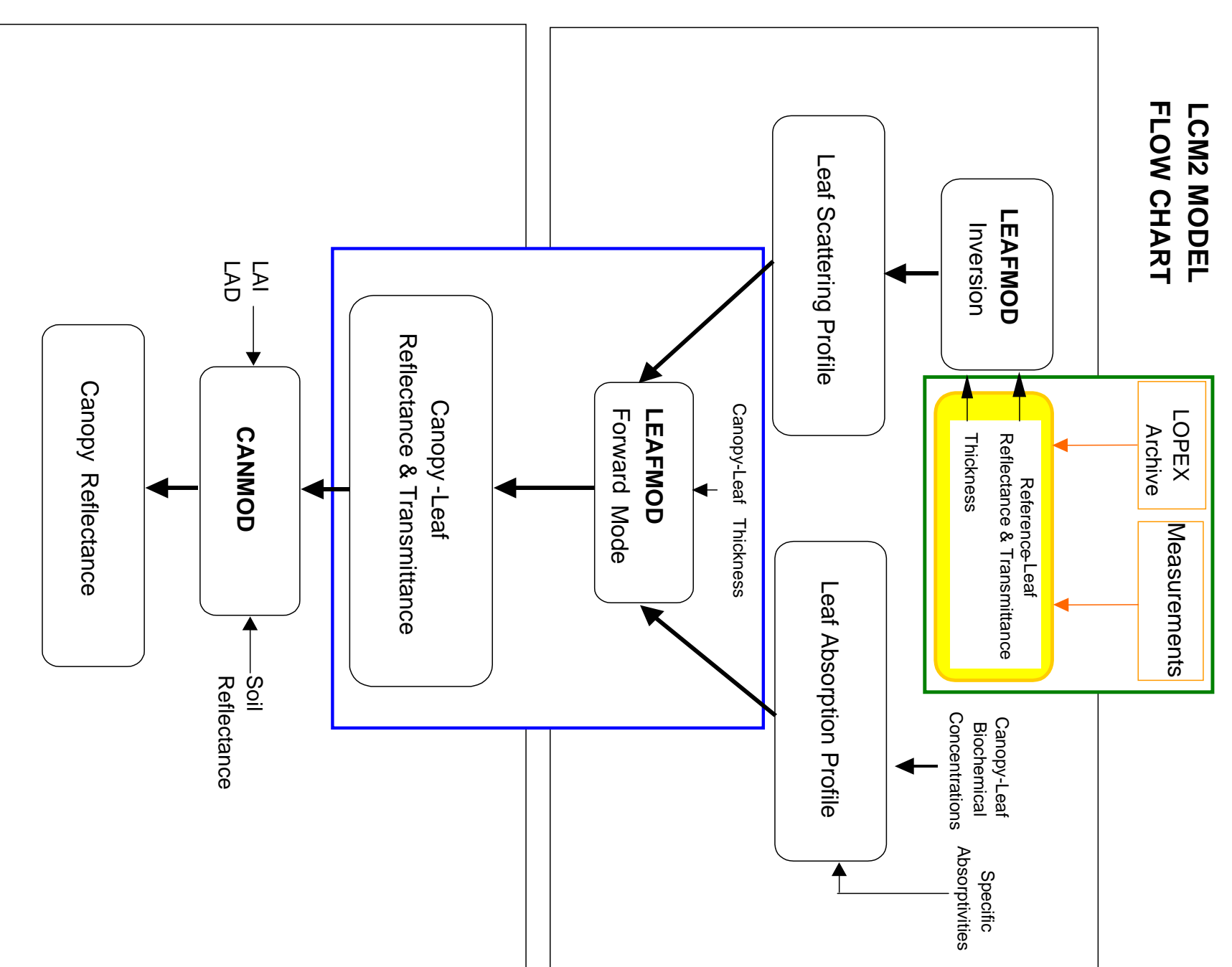


FIGURE 1: Flow chart demonstrating the operations of the coupled algorithm of the LCM

- The LAD variable is set to planophile (leaves mostly horizontal), and the sun angle is set to zenith.
- **Input variables:** Chlorophyll, water, thickness, lignin, protein, LAI, and soil reflectance.
- **Output:** $y = f(\mathbf{v})$ is hemispherical reflectance, which is the LCM output to the inputs listed above and denoted by $\mathbf{v} = (v_1, v_2, \dots, v_7)$.

Global Sensitivity Analysis

- Output function Decomposition:

$$y = f(\mathbf{v}) = E(y) + \sum_{i=1}^7 z_i(v_i) + \sum_{i < j} z_{i,j}(v_i, v_j) + \dots + z_{1,2,\dots,7}(v_1, v_2, \dots, v_7) \quad (1)$$

- The **global mean** is given by $E(y) = \int_{\mathbf{v}} f(\mathbf{v}) dH(\mathbf{v})$, where $H(\mathbf{v})$ is the distribution of the inputs. Based on related literature, we use independent uniforms over the ranges of the inputs.
- The **main effects** are given by $z_i(v_i) = E(y|v_i) - E(y) = \int_{\mathbf{v}_{-i}} f(\mathbf{v}) dH(\mathbf{v}_{-i}|v_i) - E(y)$, where \mathbf{v}_{-i} denotes all the elements of \mathbf{v} except v_i .
- The later terms of the decomposition are the interactions, which give the combined influence of two or more inputs taken together.
- Computing the main effects requires the evaluation of multidimensional integrals over the input space of the model, and evaluating RTMs can be computationally expensive.
- We use a Gaussian Process (GP) approximation to the RTM output, a technique known in statistical literature as *emulation*. This approximation allows for evaluating the main effects analytically.

Bayesian Model for the Gaussian Process

- A GP is a stochastic process that generates a collection of random variables, any finite number of which have a multivariate normal distribution.
 - Given the input $\mathbf{v} = (v_1, v_2, \dots, v_7)$, a GP, $f(\mathbf{v})$, is fully specified by its mean function, $\mu(\mathbf{v})$, and covariance function, $C(\mathbf{v}, \mathbf{v}') = \tau^2 R(\mathbf{v} - \mathbf{v}')$, which is taken to be isotropic with constant variance τ^2 and a correlation function, $R(\mathbf{v} - \mathbf{v}')$.
 - Consider the exponential correlation function:
- $$R(\mathbf{v} - \mathbf{v}') = \exp \left\{ - \sum_{j=1}^7 \phi_j |v_j - v'_j| \right\}, \quad \text{where } \phi_j > 0. \quad (2)$$
- Assume the data, $D = \{y_i, \mathbf{x}_i = (x_{i1}, x_{i2}, \dots, x_{i7}) : i = 1, 2, 3, \dots, n\}$, are a sample from a GP (i.e. we are approximating the function $y = f(\mathbf{v})$ by a GP).
 - We place priors on the parameters of the GP, μ , τ^2 , and ϕ_j , for $j = \{1, 2, \dots, 7\}$, as follows.

$$\mu \sim N(a_\mu, b_\mu) \quad \tau^2 \sim \text{InvGamma}(a_\tau, b_\tau) \quad \phi_j \sim \text{Unif}(0, b_{\phi_j}) \quad (3)$$

- Using MCMC simulation, we obtain samples of $P(\boldsymbol{\theta}, \mu, \tau^2, \boldsymbol{\phi} | D)$, which is the joint posterior distribution of μ , τ^2 , $\boldsymbol{\phi} = (\phi_1, \phi_2, \dots, \phi_7)$, and $\boldsymbol{\theta} = (f(\mathbf{x}_1), f(\mathbf{x}_2), \dots, f(\mathbf{x}_n))$.
- Calculating the main effects requires computing $E^*\{E(y|v_i) | D\}$ and $E^*\{E(y) | D\}$, where $E^*\{\cdot | D\}$ indicates the expectations with respect to the GP posterior predictive distributions.

$$E^*\{E(y) | D\} = \int \left(\mu + T^T R^{-1} (\boldsymbol{\theta} - \mu \mathbf{1}_n) \right) dP(\boldsymbol{\theta}, \mu, \tau^2, \boldsymbol{\phi} | D). \quad (4)$$

And for each value u_j of the j -th input, we have:

$$E^*\{E(y|u_j) | D\} = \int \left(\mu + T_j^T (u_j) R^{-1} (\boldsymbol{\theta} - \mu \mathbf{1}_n) \right) dP(\boldsymbol{\theta}, \mu, \tau^2, \boldsymbol{\phi} | D), \quad (5)$$

where T and $T_j(u_j)$ are $n \times 1$ vectors that are functions of the input space and of $\boldsymbol{\phi}$, and R is the observed correlation matrix.

Some Results

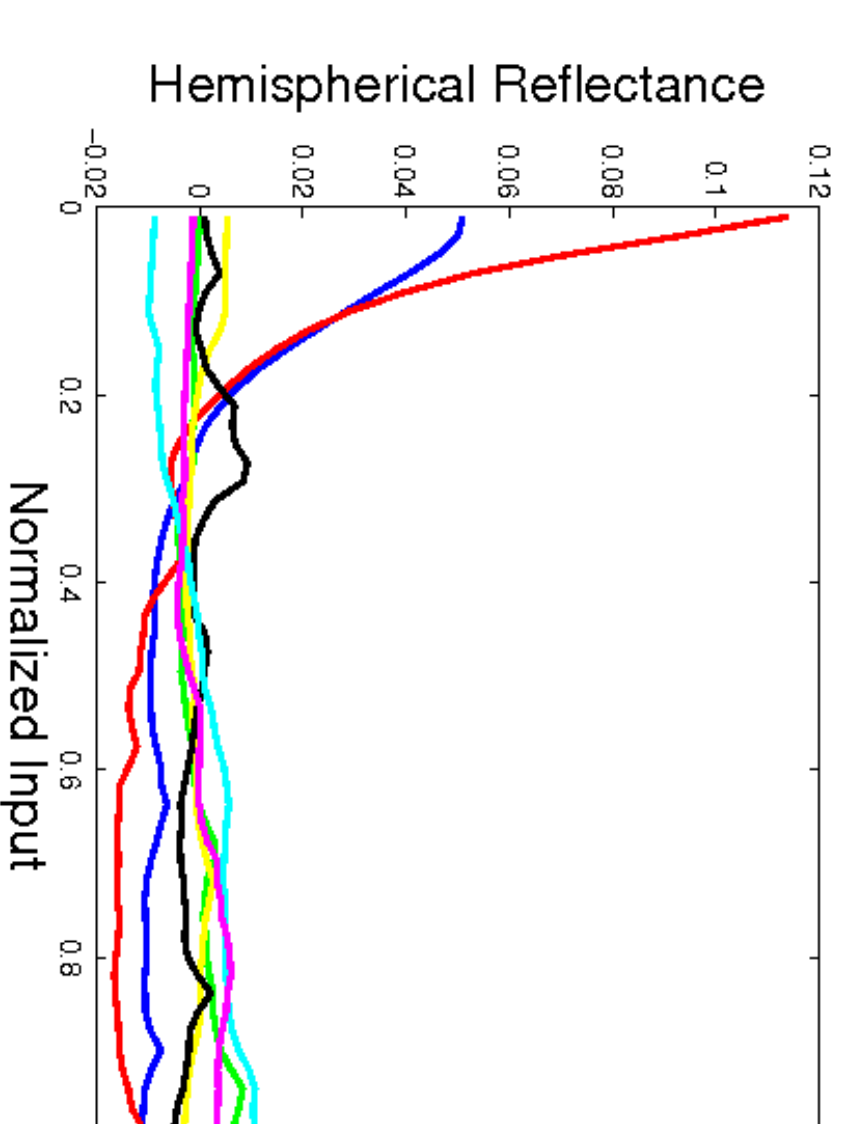


FIGURE 2: The main effects for the LCM at wavelength 667 (nm).

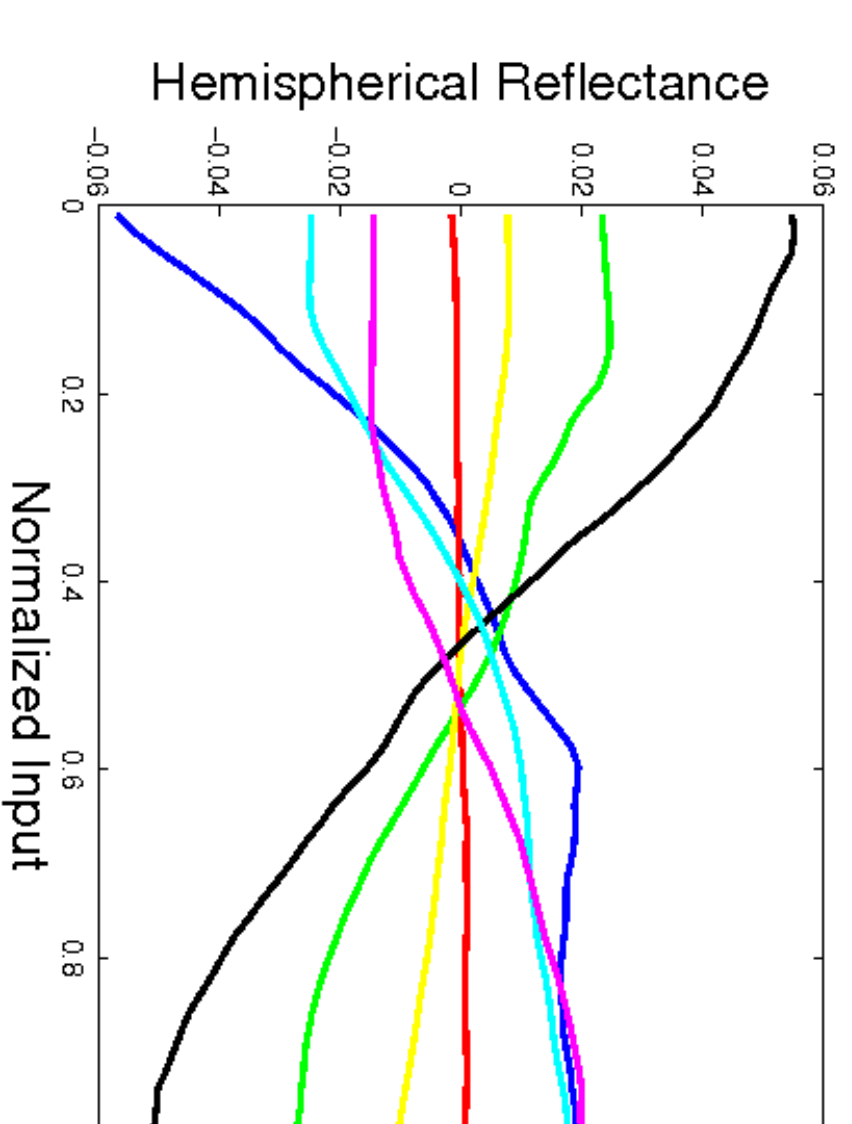


FIGURE 3: The main effects for the LCM at wavelength 1640 (nm).

- The slope of each main effect gives information as to whether the output is an increasing or decreasing function of that input.
- For example, at wavelength 667, the LCM is most sensitive to LAI and chlorophyll, which have nonlinear effects, and an increase in LAI or chlorophyll produces a decrease in reflectance.

Discussion and Future Work

- We have implemented a Bayesian approach, via MCMC methods for the GP emulator, to obtain point estimates for all the main effects associated with the 7 inputs at two different MODIS wavelengths. To quantify the uncertainty introduced by the GP approximation of the LCM, we will obtain $Var^*\{E(y|v_i) | D\}$.
- Using a fully Bayesian approach, we will obtain distributions (rather than point estimates) of the main effects associated with the 7 inputs at 8 MODIS band as well as “sensitivity indices”, which give a measure of how much of the variance of the output is due to each input.

- We will also develop a hierarchical Bayesian model accounting for all 4 LAD classifications using a hierarchical GP formulation for the corresponding output functions $f_j(\mathbf{v})$, for $j = 1, 2, 3, 4$.

Acknowledgements:

This work was supported by the NASA AISR program through grant number NNX07AY69G.

Bayesian Inference for Global Sensitivity Analysis of Radiative Transfer Models

Marian Farah, Athanasios Kottas, and Robin Morris
Department of Applied Math and Statistics, University of California, Santa Cruz



Introduction

- Global process models are widely used in geoscience and remote sensing for the estimation and prediction of the properties of Earth's coupled dynamical system. Such models are typically implemented in complex computer programs that require global inputs.
- We are concerned with Radiative Transfer Models (RTMs), which simulate light reflected off the surface of the earth. RTMs are typically computationally expensive, and while they are deterministic models, there is uncertainty about the true values of their inputs.
- We use the LeafCanopy Model (LCM) as a surrogate for the RTM, and study the sensitivity of the LCM's output to uncertainty in its inputs using global sensitivity analysis. Then, we determine inputs that are most influential with regard to the LCM output prediction uncertainty.

Leaf-Canopy Model

- The LCM was developed in support of MODIS (or Moderate Resolution Imaging Spectroradiometer), a key instrument aboard Terra and Aqua satellites, in order to capture essential biophysical processes associated with the interaction between light and vegetation.
- The LCM combines two radiative transfer algorithms: LEAFMOD, which simulates the radiative regime inside the single leaf, and CANMOD, which combines the information coming from LEAFMOD with canopy structural parameters to compute the radiative regime within and at the top of the canopy.

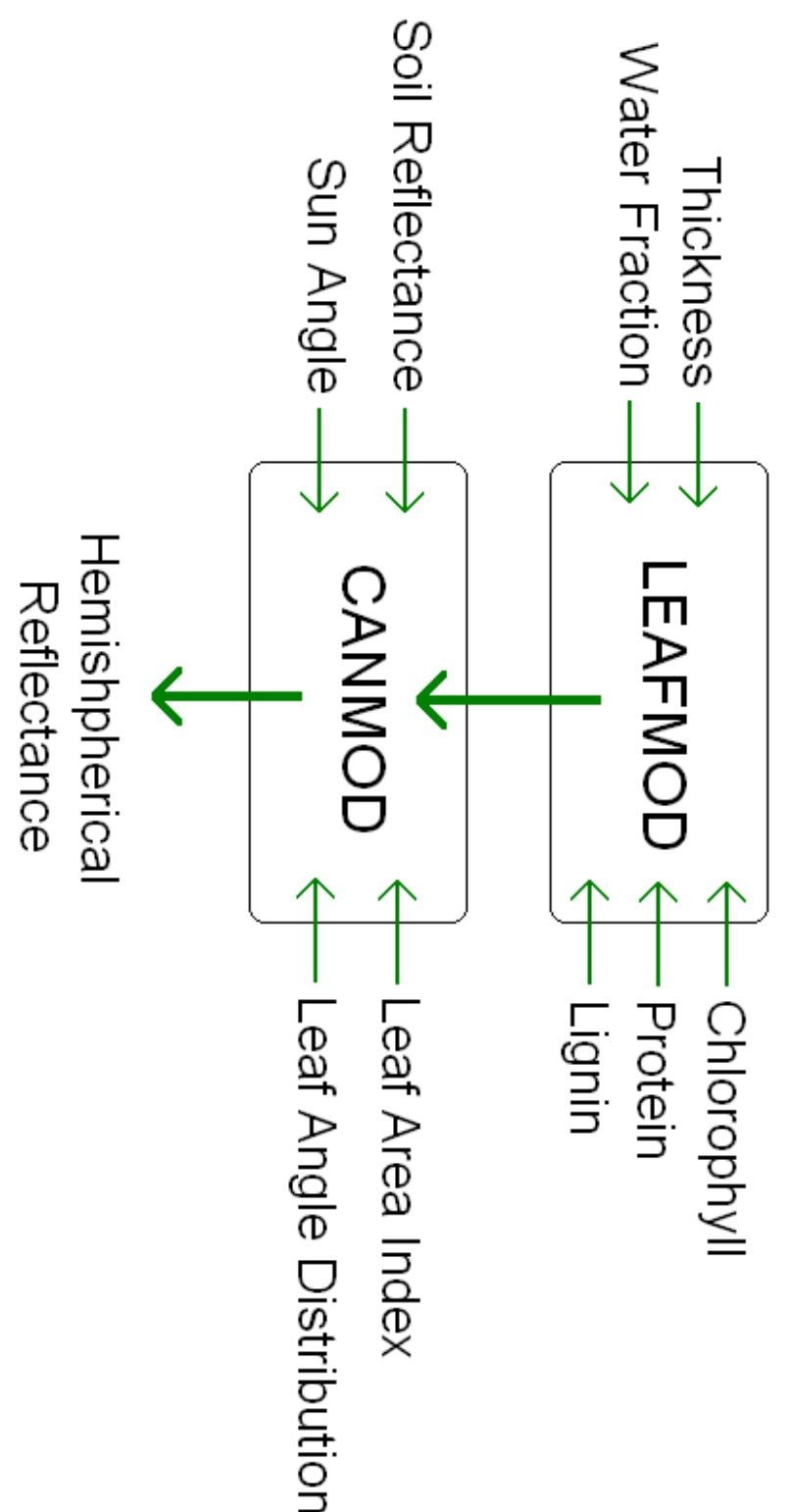


FIGURE 1: The inputs and output of the coupled algorithm of the LCM

- Input variables:** We set the leaf angle distribution to planophile (leaves mostly horizontal) and the sun angle to zenith, and consider chlorophyll, water fraction, thickness, lignin, protein, Leaf Area Index (or LAI), and soil reflectance, denoted by $\mathbf{v} = (v_1, v_2, \dots, v_7)$.
- Output:** $y = f(\mathbf{v})$ is hemispherical reflectance, which is the LCM output given inputs \mathbf{v} .

Global Sensitivity Analysis

- The influence of each input and how uncertainty in the output is apportioned amongst the inputs are determined by calculating the "main effects" and "sensitivity indices" of the LCM inputs.
- Output function Decomposition:

$$y = f(\mathbf{v}) = E(y) + \sum_{i=1}^7 f_i(v_i) + \sum_{i < j} f_{i,j}(v_i, v_j) + \dots + f_{1,2,\dots,7}(v_1, v_2, \dots, v_7)$$

- The **global mean** is given by $E(y) = \int_{\mathbf{v}} f(\mathbf{v}) dH(\mathbf{v})$, where $H(\mathbf{v})$ is the distribution of the inputs. Based on related literature, we use independent uniforms over the ranges of the inputs.
- The **main effects** are given by $f_i(v_i) = E(y|v_i) - E(y) = \int_{\mathbf{v}_{-i}} f(\mathbf{v}) dH(\mathbf{v}_{-i}|v_i) - E(y)$, where \mathbf{v}_{-i} denotes all the elements of \mathbf{v} except v_i . The latter terms of the decomposition are the interactions, which give the combined influence of two or more inputs taken together.

- Assuming independence between the input variables in the uncertainty distribution, $H(\mathbf{v})$, the total variance, $\text{Var}(Y) = W$, can be decomposed as the sum of partial variances,

$$W = \sum_{i=1}^7 W_i + \sum_{1 \leq i < j \leq 7} W_{i,j} + \dots + W_{1,2,\dots,7}.$$

$W_i = \text{Var}(f_i(v_i)) = \text{Var}\{E(Y|v_i)\}$, $W_{i,j} = \text{Var}(f_{i,j}(v_i, v_j)) = \text{Var}\{E(Y|v_i, v_j)\} - W_i - W_j$, and analogously for the higher order terms.

- The **sensitivity indices**, are given by

$$S_i = \frac{W_i}{W}, S_{i,j} = \frac{W_{i,j}}{W}, \dots, S_{1,2,\dots,7} = \frac{W_{1,2,\dots,7}}{W}.$$

where S_i is the *first-order sensitivity index*, S_{ij} , for $i \neq j$, is the *second-order sensitivity index*, and so on. We're interested in S_i which measures the fractional contribution of v_i to $\text{Var}(Y)$.

- Another important sensitivity measure is the **total sensitivity index**, $S_i^T = 1 - (W_{-i}/W)$, where W_{-i} is the total contribution to the variance of $f(\mathbf{v})$ due to all inputs except v_i .
- Computing the main effects and sensitivity indices requires the evaluation of multidimensional integrals over the input space of the model. The LCM is computationally expensive, so obtaining these quantities through Monte Carlo methods using LCM runs is not feasible.
- Using a Bayesian approach, we approximate the LCM with a Gaussian Process (GP) emulator and efficiently obtain posterior inference for the main effects and sensitivity indices.

Bayesian Gaussian Process Emulator

- A GP is a stochastic process that places a probability distribution over a function, $f(\cdot)$, such that given a finite set of input points, $\mathbf{v} = (v_1, v_2, \dots, v_k)$, the joint probability distribution of $(f(v_1), f(v_2), \dots, f(v_k))$ is multivariate normal.
- A GP is fully specified by its mean function, $\mu(\mathbf{v})$, and covariance function, $C(\mathbf{v}, \mathbf{v}') = \tau^2 R(\mathbf{v} - \mathbf{v}')$. We assume a constant mean, μ , and an isotropic covariance function with constant variance, τ^2 , and correlation function, $R(\mathbf{v} - \mathbf{v}') = \exp\left\{-\sum_{j=1}^7 \phi_j |v_j - v'_j|\right\}$, where $\phi_j > 0$.
- We use the GP to formulate a prior distribution for the function $f(\mathbf{v})$. Then, using a small number of carefully chosen RTM runs, we obtain a posterior distribution according to Bayes' Rule using Markov chain Monte Carlo (MCMC) sampling.
- The main effects and sensitivity indices of the LCM inputs are then obtained using computationally "cheap" runs of the GP posterior.

Results

- We construct the Bayesian GP emulator using a training set of 250 LCM runs based on a Latin hypercube design at 8 MODIS spectral bands that are sensitive to vegetation.

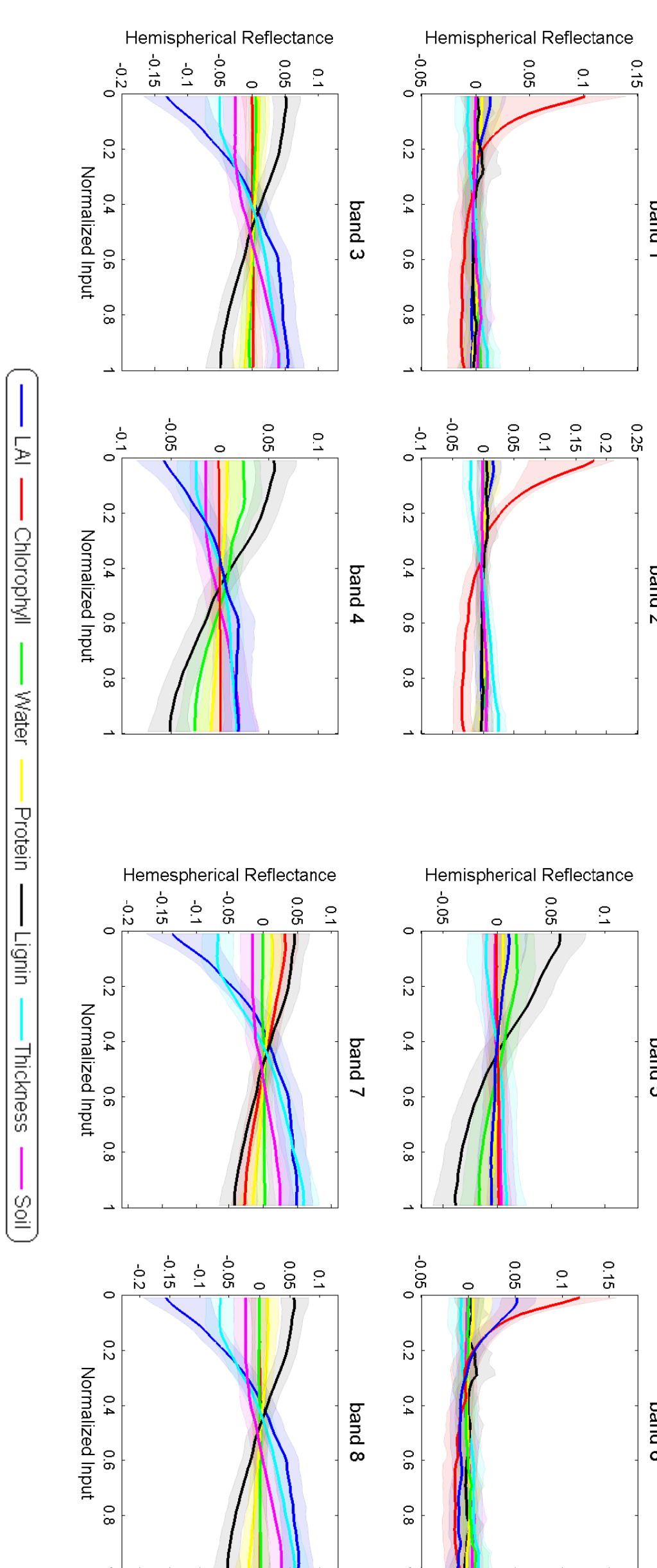


FIGURE 2: Medians (smooth lines) and 95% probability bands (the shaded regions around the medians) of the posterior distributions of main effects of the LCM at 8 different MODIS bands.

- Normalizing the inputs allows all the main effects to be plotted together on the same plot. The larger the variation of the main effect plot, the greater the influence of that input on the LCM output. The slope of each main effect plot gives information as to whether the output is an increasing or decreasing function of that input.

- For visible spectrum (bands 1, 2, & 6), the LCM is most sensitive to chlorophyll, and an increase in chlorophyll results in a decrease in the LCM output. For red light (band 6), LAI becomes important as well, and an increase in LAI results in a decrease in the LCM output.

- For near infrared (bands 3, 7, & 8), the LCM is most sensitive to LAI, lignin, and thickness (in that order), and an increase in LAI or thickness produces an increase in the LCM output, while an increase in lignin produces a decrease in the LCM output.

- For short infra-red bands (bands 4 & 5), LAI and lignin continue to be influential inputs, with water fraction also becoming influential. An increase in LAI produces an increase in the LCM output, while an increase in lignin produces a decrease in the LCM output.

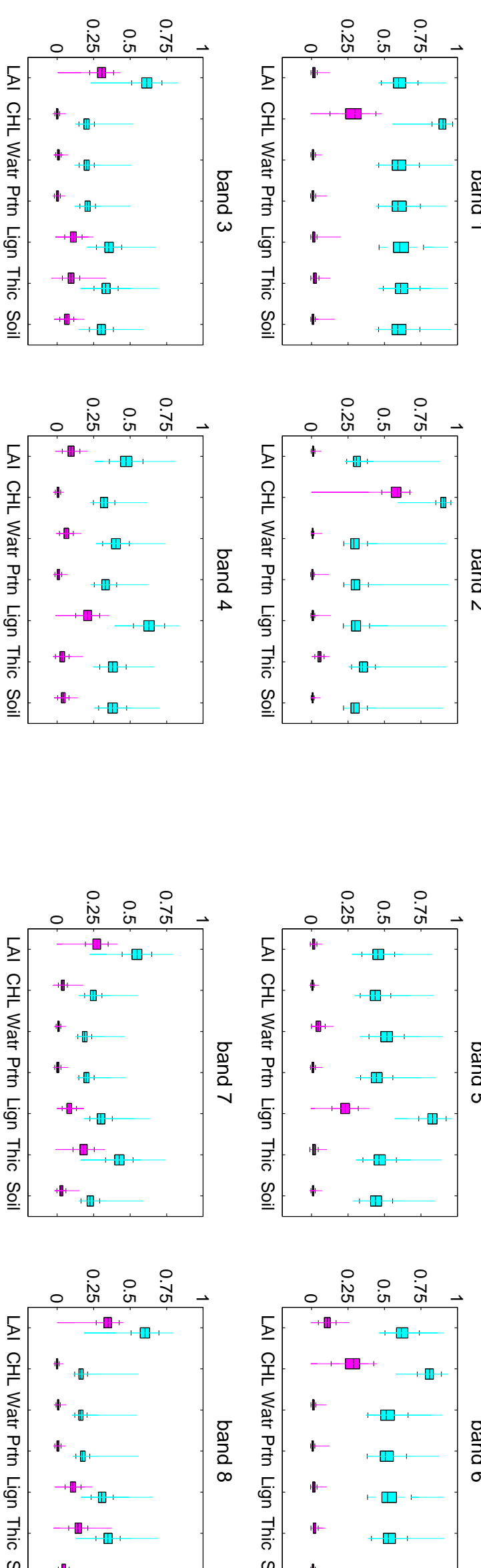


FIGURE 3: The distributions of the first-order sensitivity indices (in magenta) and the total sensitivity indices (in cyan) of the LCM inputs as estimated by the GP emulator.


- The box plots of the sensitivity indices show that inputs with influential main effects also have large sensitivity indices, which means they are major contributors to the variation in the LCM output.
- Many inputs with negligible (nearly zero) first-order sensitivity indices had non-negligible total sensitivity indices. A substantial difference between the first-order sensitivity index and total sensitivity index of a particular input implies an important role of interaction terms involving that input on the variation in the output.
- For all 8 MODIS bands, we find that interaction terms involving the 7 LCM inputs are influential in controlling output variability, which indicates that the dimension of the LCM input is irreducible.

Discussion and Future Work

- We have implemented a Bayesian approach, via MCMC methods for the GP emulator, to obtain posterior inference for the main effects and sensitivity indices associated with the 7 LCM inputs at 8 different MODIS bands.
- Our analysis enabled the identification of influential first-order effects of the inputs to the LCM and revealed that interaction terms are also important in controlling the variation of the LCM output.
- We plan to study a Bayesian variable selection approach in the context of sensitivity analysis, where the GP correlation parameters are used to make screening decisions in order to reduce the input space by identifying "active" inputs.
- The long-term goal is to validate the LCM using field data, and to invert the LCM in order to obtain the distribution of Leaf Area Index, a key input to global climate models, over large geographic regions, given measured reflectances from the satellite data.

Acknowledgements:

This work was supported in part by the NASA AISR program through grant number NNX07AV69G.

		Cover Page for Proposal Submitted to the National Aeronautics and Space Administration		NASA Proposal Number 08-OUTREACH08-0007	
NASA PROCEDURE FOR HANDLING PROPOSALS					
This proposal shall be used and disclosed for evaluation purposes only, and a copy of this Government notice shall be applied to any reproduction or abstract thereof. Any authorized restrictive notices that the submitter places on this proposal shall also be strictly complied with. Disclosure of this proposal for any reason outside the Government evaluation purposes shall be made only to the extent authorized by the Government.					
SECTION I - Proposal Information					
Principal Investigator Robin Morris			E-mail Address rdm@riacs.edu		Phone Number 650-966-5035
Street Address (1) 444 Castro St			Street Address (2) Suite 320		
City Mountain View		State / Province CA		Postal Code 94041-2017	Country Code US
Proposal Title : Living in an Uncertain World: a video for public television					
Proposed Start Date 07 / 01 / 2009		Proposed End Date 06 / 30 / 2010		Total Budget No budget required	
SECTION II - Application Information					
NASA Program Announcement Number NNH08ZDA001N-OUTREACH		NASA Program Announcement Title Supplemental Outreach Awards for ROSES Investigators			
For Consideration By NASA Organization <i>(the soliciting organization, or the organization to which an unsolicited proposal is submitted)</i> NASA , Headquarters , Science Mission Directorate , Cross Division					
Date Submitted 12 / 15 / 2008		Submission Method Electronic Submission Only		Grants.gov Application Identifier	Applicant Proposal Identifier 09-036
Type of Application New	Predecessor Award Number		Other Federal Agencies to Which Proposal Has Been Submitted		
International Participation Yes	Type of International Participation PI				
SECTION III - Submitting Organization Information					
DUNS Number 056912900	CAGE Code 1F731	Employer Identification Number (EIN or TIN) 520892064		Organization Type 8H	
Organization Name (Legal Name) UNIVERSITIES SPACE RESEARCH ASSOCIATION				Company Division	
Organization DBA Name U S R A				Division Number	
Street Address (1) 10211 WINCOPIN CIR 500			Street Address (2)		
City COLUMBIA		State / Province MD		Postal Code 210443405	Country Code USA
SECTION IV - Proposal Point of Contact Information					
Name Robin Morris		Email Address rdm@riacs.edu		Phone Number 650-966-5035	
SECTION V - Certification and Authorization					
Certification of Compliance with Applicable Executive Orders and U.S. Code					
By submitting the proposal identified in the Cover Sheet/Proposal Summary in response to this Research Announcement, the Authorizing Official of the proposing organization (or the individual proposer if there is no proposing organization) as identified below:					
<ul style="list-style-type: none"> certifies that the statements made in this proposal are true and complete to the best of his/her knowledge; agrees to accept the obligations to comply with NASA award terms and conditions if an award is made as a result of this proposal; and confirms compliance with all provisions, rules, and stipulations set forth in the two Certifications and one Assurance contained in this NRA (namely, (i) the Assurance of Compliance with the NASA Regulations Pursuant to Nondiscrimination in Federally Assisted Programs, and (ii) Certifications, Disclosures, and Assurances Regarding Lobbying and Debarment and Suspension. 					
Willful provision of false information in this proposal and/or its supporting documents, or in reports required under an ensuing award, is a criminal offense (U.S. Code, Title 18, Section 1001).					
Authorized Organizational Representative (AOR) Name edward abrams		AOR E-mail Address eabrams@hq.usra.edu		Phone Number 410-740-6231	
AOR Signature <i>(Must have AOR's original signature. Do not sign "for" AOR.)</i>				Date	

PI Name : Robin Morris		NASA Proposal Number	
Organization Name : UNIVERSITIES SPACE RESEARCH ASSOCIATION		08-OUTREACH08-0007	
		NASA Proposal Number	
Proposal Title : Living in an Uncertain World: a video for public television			
SECTION VI - Team Members			
Team Member Name Robin Morris		E-mail Address rdm@riacs.edu	Phone Number 650-966-5035
Organization Name USRA-RIACS		Team Member Role PI	International Participation No
U.S. Government Agency Participation No	U.S. Government Agency		Total Funds Requested 0.00
Team Member Name Brett Casadonte		E-mail Address brett@casadonte.net	Phone Number 650-483-7713
Organization Name The Casadonte Group LLC		Team Member Role Co-I	International Participation No
U.S. Government Agency Participation No	U.S. Government Agency		Total Funds Requested 0.00

PI Name : Robin Morris	NASA Proposal Number
Organization Name : UNIVERSITIES SPACE RESEARCH ASSOCIATION	08-OUTREACH08-0007
	NASA Proposal Number

Proposal Title : **Living in an Uncertain World: a video for public television**

SECTION VII - Project Summary

We will develop a five minute video to explain uncertainty estimation in Earth Science data and models. The video will give examples to show how optimal decisions change depending on the level of uncertainty about the issues. Examples will be given from the Earth Science domain of the parent proposal, estimating the uncertainty of remote sensed data products. The video will be broadcast several times over the coming year on a local PBS station between regularly scheduled science programming such as NOVA. A Spanish version of the video will also be developed, and additional distribution of the videos will be done via YouTube and the proposing institutions web sites. The video is intended to generate large scale awareness of the important research being conducted in modeling the Earth's ecosystem, and inform people of how probability-based models are allowing a better understanding of the world we live in. Whilst not being curriculum-aligned, the video will have significant educational content.

Contents

1	Scientific/Technical/Management Section	12
1.1	Introduction	12
	Audience	12
	Access and Diversity	13
	Leadership and Team	13
1.2	Methodology	13
	Pipeline	14
1.3	Impact	14
1.4	Evaluation	15
1.5	Work Plan and Key Milestones	15
2	References	16
2.1	List of Acronyms	16
3	Curriculum Vitae	17
3.1	Dr Robin D. Morris	17
3.3	Brett Casadonte	19
4	Current and Pending Support	20
5	Letters of Commitment	21
6	Letters of Support	22
7	Budget	23
7.1	Budget Table	23
7.2	Summary of Proposal Personnel and Work Effort	23
7.3	Facilities and Equipment	23
7.4	Budget Narrative	24
7.5	In-kind Contributions	24
8	Budget Details	25
8.1	Summary of Cost and Fee	25

1.1 Introduction

Communicating the excitement generated by scientific research is a vital part of NASA's mission. Education and Public Outreach is a key agency priority, sufficiently important to be part of the Strategic Management Framework in the 2006 NASA Strategic Plan [1, pp 29-30], where NASA's Education Initiatives are outlined. These include "Engage[ing] Americans in NASA's mission", which becomes Education Outcome ED-3 [1, p43], "... promote STEM literacy and awareness of NASA's mission".

To further these NASA goals we propose to produce a five minute video on uncertainty estimation, to be aired on local PBS educational affiliate KSCM-TV (3.5 million viewers) as well as YouTube and other venues. The video will present uncertainty estimation in non-mathematical terms in order to inform the viewer about how uncertainty is estimated, how it is used in scientific decision making, and how optimal decisions change depending on the level of uncertainty. Examples will be given from the Earth Science domain of the parent proposal, estimating the uncertainty of remote sensed data products from NASA's Earth Observing satellites.

The proposed video, **"Living in and Uncertain World"**, will focus on how uncertain knowledge can impact decision making, and will provide simple examples comprehensible to a high-school educated person, before showing how the same ideas can be used to improve the data products NASA produces from remote sensed imagery. The video will directly involve NASA science, in a focused way, in terms of the explanation to a scientifically literate lay person of the principles of uncertainty; how uncertainty often has a different meaning when used in the scientific domain as opposed to general conversation; and how knowledge of the uncertainty associated with a measurement can affect decisions. The video will also draw in more general NASA science, through the use of NASA remote sensing data as examples of where the uncertainty modeling is being applied. This is of great current public interest, as many of the debates about the causes and effects of climate change rest on the uncertainty in the measurements and models.

To present these topics to the viewer, the video will integrate interviews conducted with experts in the field of uncertainty with motion graphics and supporting video footage in order to enhance visual impact and convey the concepts explained in the video. The inclusion of scientists allows them to be seen as role models, and the participating scientists will be chosen to include women and minority researchers, to represent the diverse spectrum of people working in SMD science. The videos will also be produced with English and Spanish subtitles, enabling outreach to a more diverse and often under-served community.

Audience

We know that there is a sizable segment of the American television viewing audience that is interested in science and scientific topics by virtue of the fact that television stations air science related programming (e.g., NOVA, Nature, etc.). This audience may have little formal science background, but are intelligent, interested in science, and receptive to informal presentation of scientific ideas. Through our collaboration with KSCM-TV, we will be able to reach this audience by developing a short-format 'filler' video to be aired between regularly scheduled science programming. For example, at the conclusion of an episode of NOVA, KSCM-TV will be able to air the proposed video as a 'lead in' to its next regularly scheduled program, exposing an interested audi-

ence to informational science programming based on NASA research. This serves two purposes: 1) to continue to provide high-quality, scientific content to demanding viewers that have already tuned-in to see science-related programming and 2) to educate an audience interested in science on advanced scientific concepts such as uncertainty and scientific decision making.

For KCSM-TV, having high-quality science content to ‘fill the gaps’ between regularly scheduled science programming is very important to maintaining viewer interest. As an educational PBS affiliate, KCSM-TV strives to provide high-quality educational content to its viewers by airing well known programs such as NOVA and Nature. However, each episode of these popular programs typically runs 52-54 minutes, leaving a 6-8 minute window between programs that KCSM-TV needs to fill with supplemental programming. KCSM-TV, like all PBS affiliates, is in need of high-quality, informational content to fill this programming gap, and more specifically, they need content that is generally aligned with the programming that surrounds it (eg, science filler for science programming.) Subsequently, high-quality ‘filler’ content is in high demand, particularly content that is scientific in nature. Currently, since there is a dearth of scientific filler content; KCSM-TV and many other PBS stations fill this gap with a hodge-podge of miscellaneous content which in many cases is not even scientific in nature.

Access and Diversity

In addition to developing the video in English, a Spanish language version will also be produced. This will further broaden access to the video, allowing the content to be seen by a traditionally under-served demographic. As with the English version, the Spanish version will be made available to KCSM-TV, and will also be distributed via YouTube as well as the USRA web site. Both English and Spanish versions will be produced with closed-caption subtitles.

Leadership and Team

The science PI, Dr Robin Morris, will be responsible for the scientific content of the script for the video, and will oversee the narration and the production, to ensure scientific accuracy in the result. The actual production will be contracted to The Casadonte Group, who have extensive experience in public outreach, video production and content development. They have produced a number of videos for USRA-RIACS that have already been broadcast on public television.

1.2 Methodology

“Living in an Uncertain World” will expose the viewer to the concepts of uncertainty and uncertainty estimation, and how knowledge of uncertainty associated with a scientific measurement can affect decisions.

To do this, we will use a series of interviews conducted with authorities in field of uncertainty estimation, asking them such questions such as, What is uncertainty? How do you estimate it? How does uncertainty affect decision making? Why can uncertainty estimation can be counter intuitive? What is Bayes Theorem? and How is it used in uncertainty estimation? Interviewees will be asked to answer these questions within the context of NASA’s Earth Science research and how uncertainty estimation is being applied to NASA remote sensing data. These concepts will be further illustrated using motion graphics, b-roll, and publicly available data from NASA’s EOS missions.

The proposers have already produced a successful and well-received video about the Fermi (GLAST) mission using this format [2].

Pipeline

The video will be broadcast on public television, and will facilitate viewers becoming aware of the role uncertainty plays in scientific decision making and how it impacts decisions made. Seeing the practical application of real scientific principles should help increase public awareness and overall understanding of science, and will hopefully lead to further inquiry by the viewer. If deemed appropriate, links to additional information on existing NASA web sites can be included in the video, to make further exploration of this topic easier for viewers.

Additionally, including researchers, particularly women and minority researchers, allows them to be seen as role models and showcase the diverse spectrum of people working in SMD science.

1.3 Impact

The proposal team has worked together in the past to produce successful videos like the one proposed here. Most recently, the team produced a video on the NASA GLAST (now Fermi) mission [2], which was released in June-July 2008. Since then, the video has and been aired nearly a dozen times on KCSM-TV, filling such coveted time slots as the gap between NOVA and following programs, and has been viewed nearly 2,000 times on YouTube. In addition, portions of the video have also been used by the GLAST team in various outreach and informal education activities.

In addition, Mr Casadonte of TCG has been involved in the production of a number of filler pieces for NASA and USRA which have been aired over 200 times by KCSM-TV over the past four years. These videos have won over a dozen Telly, Aurora, and Beacon awards for excellence in video production. This video continues to build upon this successful experience by adding additional scientific content to an established and broad-reaching collaboration.

KCSM-TV will be the primary distribution channel for the videos, which will be aired multiple times over a two to three year period, ensuring good re-use of the content. (It is not uncommon for videos to be placed in the rotation schedule for several months, removed from the rotation for a short period, and then placed back into rotation in order to appear ‘fresh’ and reach new viewers.) Considering the current lack of high-quality scientific fillers, it is anticipated that these videos will be in circulation for many years beyond the initial time period.

In addition to KCSM-TV, the video will be distributed on YouTube as well as USRA’s web sites, www.usra.edu and www.riacs.edu. The Spanish language version will be made available through these distribution methods as well.

By tapping into the airing opportunities available through KCSM-TV, this video provides access to 3.5 million potential viewers in the San Francisco Bay Area, and can serve as a demonstration of what can be accomplished via this distribution channel, potentially leading to a larger EPO project in the future with national PBS coverage.

1.4 Evaluation

To ensure that the video's content is suitable for the audience, and that it clearly explains the concepts we wish to communicate, we will work closely with our partners at KCSM-TV when developing the storyboard, soliciting their feedback. We will present the storyboard to a small number of people selected to be representative of the PBS audience. We will use the feedback from these presentations to modify the storyboard to ensure that the intended audience understands the video's content. This feedback will allow us to confirm that any points of confusion have been eliminated and that the final video will be clear and informative.

A separate, small focus group will review and evaluate the first rough cut of the video. This group, which will be different from the persons who reviewed the storyboard, will provide any additional feedback on the video, and from this feedback, final edits will be made.

While broad evaluation of the longitudinal impact of the video is not possible given the budget constraints of this proposal, we can approximate the reach of the video through access to carriage reports from KCSM-TV. A monthly report can be obtained from KCSM-TV to show when the video aired on television, and the television viewership around the time the production was aired. This will provide an indication of the size of the audience for the video, and hence its potential impact.

1.5 Work Plan and Key Milestones

The following production timeline is offered based on past experience developing video projects of the scope described in this proposal:

Weeks 1, 2 and 3: Storyboarding, Pre-production, and Scheduling During this phase, the storyboard for the video will be fully developed and a pre-evaluation will be conducted. Based on feedback from the pre-evaluation, we will revise the storyboard and schedule interviews. At this time, the following people are anticipated as being our most likely interviewees: Professors Raquel Prado and Bruno Sanso, University of California, Santa Cruz; Drs. Jennifer Dungan, Joe Coughlan and Rama Nemani, NASA Ames.

Week 4: Production During this time we will conduct interviews with persons identified during the pre-production phase of the project. Since all of the interviewees will be local, we will be able to keep scheduling efficient and conduct the interviews in a short period of time.

Weeks 5 and 6: Post-Production During this time, the first rough-cut of the video will be produced. It will then be reviewed and screened to a small focus group to get feedback on how well. Based on this feedback, revisions to the video will be made and the final video will be produced. After the final, English language version of the video is complete a version with Spanish subtitles added will be assembled. To ensure accurate scientific translation, we will work with Spanish-speaking faculty in the statistics department at the University of California, Santa Cruz.

Milestone: Delivery completed video.

Delivery of completed video to KCSM-TV for local airing and national distribution; compression of the videos and posting on YouTube and USRA web sites.

2 References

References

- [1] 2006 NASA Strategic Plan, NP-2006-02-423-HQ http://nasascience.nasa.gov/about-us/science-strategy/142303main_2006_NASA_Strategic_Plan_sm.pdf
- [2] “NASA’s GLAST Mission”, <http://youtube.com/watch?v=3e6RvgBQMvU>

2.1 List of Acronyms

GLAST	Gamma-ray Large Area Space Telescope
PBS	Public Broadcasting System
RIACS	Research Institute for Advanced Computer Science
SMD	Science Mission Directorate
STEM	Science, Technology, Engineering, and Mathematics
TCG	The Casadonte Group
UCSC	University of California, Santa Cruz
USRA	Universities’ Space Research Association